

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

Єфремов Костянтин Вікторович

УДК 004.75

**ПЛАТФОРМА ПІДТРИМКИ МІЖДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНЬ
В СВІТОВІЙ СИСТЕМІ ДАНИХ**

Спеціальність: 05.13.06 – Інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2021

Дисертацією є рукопис.

Роботу виконано на кафедрі математичних методів системного аналізу Інституту прикладного системного аналізу та в навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор,
академік Національної академії наук України,
заслужений діяч науки і техніки України
Згуровський Михайло Захарович,
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря
Сікорського»,
ректор

Офіційні опоненти: доктор фізико-математичних наук, професор,
заслужений діяч науки і техніки України
Глибовець Микола Миколайович
Національний університет «Києво-Могилянська
академія»,
професор кафедри інформатики

доктор технічних наук, професор,
Снитюк Віталій Євгенович
Київський національний університет імені Тараса
Шевченка,
декан факультету інформаційних технологій

Захист відбудеться «28» квітня 2021 р. о 15:00 годині на засіданні спеціалізованої вченої ради Д 26.002.29 у Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, проспект Перемоги 37, корпус № 11, аудиторія № 215.

З дисертацією можна ознайомитися у Науково-технічній бібліотеці ім. Г. І. Денисенка Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, просп. Перемоги 37.

Автореферат розісланий « 26 » березня 2021 р.

Учений секретар
спеціалізованої вченої ради

С. Ф. Теленик

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Стрімкий розвиток інформаційних технологій (ІТ) створив умови для розподілених обчислень, ефективного використання інформаційних та інших ресурсів. Консолідація ресурсів, їх віртуалізація, сервісний підхід приводять до заміни локальних рішень розподіленими, які дозволяють комплексно використовувати всі об'єднані в глобальну мережу обчислювальні потужності, системи збереження даних і забезпечують доступ до накопичених інформаційних ресурсів.

Так, один із потужних інформаційних ресурсів використовує створена в 1956 р. під егідою Міжнародної наукової ради (International Science Council – ISC, <https://council.science/>) Світова система даних (World Data System – WDS, ССД, <https://www.worlddatasystem.org/>), яка забезпечує збір, збереження, поширення і аналіз даних, отриманих в багатьох галузях науки. Накопичені в цій системі дані та застосунки можна використовувати для вирішення актуальних проблем суспільства. Цьому перешкоджають труднощі у взаємодії успадкованих застосунків, використанні даних із різних джерел, формуванні результатів опрацювання запитів, що вимагають аналітичного опрацювання даних із різних джерел, застосування наявних і створення нових застосунків, обумовлені відмінністю архітектур, різноманіттям форматів представлення даних та іншими чинниками. Виникає потреба у платформі підтримки міждисциплінарних досліджень в Світовій системі даних, здатній забезпечити інтеграцію застосунків і ресурсів із різних джерел, виконання комплексного аналітичного оброблення даних, при якій вчені різних галузей знання зможуть використовувати накопичені ресурси, орієнтуючись на знання предметної області, а не особливості ІТ.

Дослідженнями, пов'язаними зі створенням такої платформи займаються низка міжнародних організацій (WDS, ISC, CODATA, Research Data Alliance), провідні університети (Гарвард, Кембрідж, Кайзерслаутерн, Берклі та ін.), а також такі провідні вчені, як Михайло Згуровський, Олександр Павлов, Євген Бодянський, Олексій Новіков, Петро Бідюк, Анатолій Каргін, Микола Глибовець, Леонід Гуляницький, Наталя Шаховська, Сергій Кривий, Володимир Опанасенко, Сергій Погорілий, Олександр Ролік, Трейсі Кімбрел, Олександр Клемм, Верн Паксон та ін. Однак, не зважаючи на значні досягнення відсутні єдине системне бачення усіх аспектів проблеми, сутності та призначення інструментарію міждисциплінарних досліджень і його місця в ССД, цілісна концепція платформи підтримки міждисциплінарних досліджень з урахуванням всього комплексу взаємозв'язків і взаємовпливів пов'язаних з нею процесів. Постає задача організації, планування і здійснення підтримки міждисциплінарних досліджень в ССД. Різний ступінь глибини опрацювання окремих аспектів інтеграції обумовив потребу в ефективних моделях, методах і засобах, необхідних для реалізації зазначених рішень. Насамперед, мова йде про математичні моделі та методи інтеграції застосунків і джерел даних, узгодження даних, впровадження проблемно-орієнтованого апарату інтелектуального аналізу даних та швидкого створення інформаційних ресурсів.

Зазначені питання є актуальними та становлять сутність досліджень цієї дисертаційної роботи.

Зв'язок роботи з науковими програмами, планами та темами. Дисертаційне дослідження виконано в рамках наукової тематики кафедри математичних методів системного аналізу Інституту прикладного системного аналізу та навчально-наукового комплексу «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського.

Робота виконувалась в рамках таких держбюджетних і договірних науково-дослідних робіт (НДР) та проектів: НДР «Розвиток комплексу баз даних та алгоритмів обробки з метою системного прогнозування поведінки складних антропогенних та природних систем» (2008-2009 рр., РК №0108U002348, НАН України); НДР «Розвиток мережі Світових центрів даних для дослідження фундаментальних основ глобального моделювання складних природних та антропогенних систем» (2009-2010 рр., РК №0109U007160, НАН України); НДР «Розробка фундаментальних основ та методів аналізу міждисциплінарних даних для створення системи інтегрованого доступу до інформаційних ресурсів Світових центрів даних Росії та України» (2010-2011 рр., РК №0110U003744, НАН України); НДР «Розробка загального підходу та методів системного узгодження даних різної природи в інфраструктурі розподілених багатодисциплінарних баз даних Російсько-Українського сегменту Світової системи даних для вирішення фундаментальних міждисциплінарних задач взаємозв'язку процесів в системі геосфер» (2012-2013 рр., РК №0113U005058, НАН України); НДР «Розробка методології та математичних основ кількісного оцінювання процесів сталого розвитку та впливу на них сукупності глобальних загроз у глобальному та регіональному контекстах» (2011-2013 рр., РК №0111U001738, МОН України); НДР «Розробка інструментарію для побудови складних багатокomпонентних моделей» (2012-2013 рр., РК №0112U000557, МОН України); НДР «Розробка інструментарію для аналізу сталого розвитку муніципалітетів» (2014-2015 рр., РК №0114U000671, НАН України); НДР «Методологія та інструментарій стратегічної екологічної оцінки проектів розвитку регіонів України», 2015-2016 рр., РК №0115U001968, РК №0116U006141, ДФФД України); НДР на замовлення підприємств і організацій України «Розробка та впровадження інформаційно-технічних засобів для збирання, обміну та спільного використання даних в галузі зняття з експлуатації атомних енергоблоків, поводження з радіоактивними відходами, ядерної та радіаційної безпеки, перетворення об'єкта «Укриття» в екологічно безпечну систему, реабілітації забруднених територій» (2016 р., ДСП «Чорнобильська АЕС»); проект «Розробка інформаційних технологій для моделювання, кількісного оцінювання та прогнозування впливу загроз виникнення конфліктів та розповсюдження зброї на стабільність розвитку суспільства в регіональному та глобальному масштабах» (2016-2017 рр, проект №6166, Український науково-технологічний центр (STCU); НДР на замовлення підприємств і організацій України «Форсайт та побудова стратегії соціально-економічного розвитку України на середньостроковому (до 2020 року) і довгостроковому (до 2030 року) часових горизонтах» (2016 р., Кооператив «Перший національний аграрний кооператив», ГО «Фундація Аграрна наддержава»); НДР «Розробка он-лайн платформи аналізу і сценарного планування сталого розвитку регіонів України в контексті якості та безпеки життя людей» (2017-2019 рр., РК №0117U002476, МОН України); НДР «Інтегрована платформа для оцінювання та сценарного планування сталого

розвитку об'єднаних територіальних громад в ході проведення адміністративно-територіальної реформи в Україні» (2020-2022 рр., РК 0120U102060, МОН України).

Мета і задачі дослідження. *Метою роботи є підвищення ефективності використання інформаційних та програмних ресурсів Світової системи даних за рахунок розроблення моделей, методів та інструментів інтеграції застосунків і джерел даних, узгодження та інтелектуального оброблення даних різної природи та створення на їх основі платформи підтримки міждисциплінарних досліджень. Основні завдання, що забезпечують досягнення мети роботи:*

1. Дослідження проблеми інтеграції застосунків і джерел даних ССД.
2. Розроблення концепції платформи підтримки міждисциплінарних досліджень в ССД.
3. Розроблення комплексу моделей і методів інтеграції застосунків ССД.
4. Розроблення комплексу моделей і методів інтеграції джерел даних ССД.
5. Розроблення механізму генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних.
6. Розроблення комплексу моделей і методів системного узгодження даних і впровадження проблемно-орієнтованого апарату інтелектуального оброблення даних.
7. Розроблення інструментальних засобів платформи підтримки міждисциплінарних досліджень в ССД.
8. Проведення експериментальних досліджень платформи підтримки міждисциплінарних досліджень в ССД.

Об'єктом дослідження є процеси підтримки міждисциплінарних досліджень Світової системи даних.

Предмет дослідження — математичні моделі, методи та інструменти інтеграції застосунків і джерел даних, а також методи узгодження та інтелектуального оброблення даних різної природи.

Методи досліджень — методи загальної теорії систем, теорії множин, формально-аналітичного моделювання, методи математичного програмування, дискретної математики та теорії штучного інтелекту, методи теорії імовірності і математичної статистики, системного аналізу, моделі та методи виведення математичної логіки.

Наукова новизна одержаних результатів:

1. Вперше створено платформу підтримки міждисциплінарних досліджень в Світовій системі даних, яка забезпечує інтеграцію застосунків і джерел даних, оброблення запитів до розподілених джерел, системне узгодження даних різної природи, проблемно-орієнтоване інтелектуальне оброблення даних та автоматизоване генерування застосунків;

2. Вперше створено метод узгодження даних, який відрізняється застосуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами;

3. Модифіковано клаузальну логіку планування взаємодії застосунків для розв'язання проблем міждисциплінарних досліджень, яка відрізняється від відомих використанням системи типів, орієнтованих на організацію взаємодії застосунків,

комплексом спеціальних аксіом описання застосунків на основі передумов та постумов і правил виведення, які визначають принципи комплексування застосунків, що дозволяє у зручному для користувача вигляді визначати проблему і будувати схему її вирішення.

4. Отримав подальший розвиток метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в просторі типів з наступною деталізацією виводу в системі індивідних об'єктів та відновленням схеми розв'язання проблеми на основі виводу.

Практичне значення одержаних результатів. Одержані в дисертаційній роботі результати можуть бути використані при розробленні та впровадженні інформаційних систем підтримки міждисциплінарних досліджень

Практичне значення мають логічні моделі опису застосунків та джерел даних, визначення семантики предметної області, перетворення даних, агента та міжагентної взаємодії при розв'язанні проблем інтеграції даних та застосунків, модель і метод генерування проблемно-орієнтованих застосунків на основі інтеграції мікросервісів, методи планування та підтримки співпраці агентів, методи та алгоритми виведення, відновлення дерева виконання запиту, федерування, побудови запитів, побудови плану виконання запиту із врахуванням коефіцієнтів ємності джерел даних, побудови зв'язаної структури даних із онтологій та підключень до джерел даних, об'єднання результатів, метод узгодження даних з врахуванням інформаційних втрат, методичні рекомендації щодо створення та налаштування проблемно-орієнтованих застосунків, які можуть бути використані в інформаційних системах, які потребують інтеграції джерел даних та застосунків, узгодження даних різної природи, проблемно-орієнтованого інтелектуального аналізу даних, швидкого розроблення і публікації інформаційних ресурсів.

Компоненти розробленої платформи підтримки міждисциплінарних досліджень в Світовій системі даних знайшли впровадження в Інституті проблем реєстрації інформації НАН України, Інституті геофізики НАН України ім. С.І. Субботіна, Інформаційно-аналітичному ситуаційному центрі КПІ ім. Ігоря Сікорського, Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського, Корпорації «Науковий парк «Київська Політехніка» та інших організаціях, що дозволило підтвердити їх працездатність, скоротити проектування і реалізацію застосунків, скоротити втрати внаслідок невчасного отримання потрібних даних, скоротити інформаційні втрати при перетворенні даних, підвищити обсяги наданих користувачам інформаційних ресурсів та підвищити ефективність аналізу даних різної природи експертами без знання інструментальних мов програмування.

Особистий внесок здобувача. Усі результати, що виносяться до захисту, отримано автором особисто. У наукових працях, опублікованих у співавторстві, з питань, що стосуються даного дослідження, здобувачу належать: логічний формалізм на основі клаузальної логіки першого порядку для планування взаємодії застосунків, метод виведення і механізм відновлення схеми розв'язання проблеми на основі виводу [8, 9]; архітектура системи інтеграції застосунків [14] та підхід до інтеграції сервісів програмних систем в Світовій системі даних для підтримки міждисциплінарних досліджень, логічні моделі опису сервісів (застосунків) та джерел

даних [17]; метод узгодження даних з врахуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами [7]; проблемно-орієнтована графічна мова аналітичної обробки даних як основа швидкого розроблення застосунків на платформі підтримки міждисциплінарних досліджень [11].

Автором розроблено платформу підтримки міждисциплінарних досліджень в Світовій системі даних та реалізовано засобами платформи тематичні застосунки, мікросервіси, віджети та сценарії (сукупність взаємопов'язаних процесів) обробки даних для вирішення низки прикладних задач [1-5, 10, 13, 15, 16].

Апробація результатів роботи. Основні положення і результати дисертаційної роботи були представлені і одержали схвалення на міжнародних науково-практичних конференціях та форумах, зокрема: International Conference on Data Sharing and Integration for Global Sustainability (Нью-Делі, Індія, 2014); Міжнародна науково-технічна конференція «Системний аналіз та інформаційні технології» (Київ, 2014, 2015, 2018); 6th Digital Earth Summit “Digital Earth in the Era of Big Data” (Пекін, Китай, 2016); ISC World Data System Members' Forum (Денвер, США, 2016); International Data Week (Денвер, США, 2016); IEEE First Ukraine Conference on Electrical and Computer Engineering (Київ, 2017); IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (Бухарест, Румунія, 2017); IEEE International Conference on System Analysis & Intelligent Computing (Київ, 2018); 14th Federated Conference on Computer Science and Information Systems (Лейпціг, Німеччина, 2019); IEEE 2nd International Conference on System Analysis & Intelligent Computing (Київ, 2020); BSUN e-Conference on “Addressing the Impact of SARS CoV2 Pandemics on the Higher Education Activities in the Black Sea Region” (Константа, Румунія, 2020); ISC World Data System Members' Forum 2020 (Токіо, Японія, 2020), », а також на наукових семінарах в КПІ ім. Ігоря Сікорського.

Публікації. За темою дисертаційної роботи опубліковано 17 наукових праць, у тому числі 5 монографій, 7 статей у наукових фахових виданнях (з них 1 стаття у періодичному науковому виданні іншої держави, яка входить до ОЕСР та/або Європейського Союзу, 2 у фахових виданнях України категорії «А»), 5 публікацій в збірниках матеріалів конференцій.

Структура та обсяг дисертації. Дисертація складається з переліку умовних скорочень, вступу, чотирьох розділів основного змісту, висновків, списку використаних джерел із 154 найменувань та 3 додатків. Загальний обсяг роботи складає 240 сторінок, її основний зміст викладено на 195 сторінках. Дисертаційна робота містить 55 рисунків та 10 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми дисертації, сформульовано мету і завдання дослідження, наукову новизну і практичне значення отриманих результатів, наведено дані про впровадження результатів роботи, апробацію результатів роботи і публікації, особистий внесок автора.

Перший розділ присвячено аналізу проблем підтримки міждисциплінарних досліджень ССД, огляду існуючих рішень та постановці задачі дослідження. На основі аналізу створення і функціонування Українського сегменту ССД та його взаємодії з іншими центрами ССД встановлено міждисциплінарний характер їх діяльності, залежність результатів досліджень від взаємодії учасників на рівні засобів обміну і перетворення даних, технологій їх оброблення і аналізу, включаючи інструменти системного узгодження міждисциплінарних даних, їх систематизації, інтелектуального оброблення, оцінювання адекватності, якості, коректності. Виконаний аналіз викрив причини неефективного використання накопичених в ССД даних і підкреслив необхідність побудови платформи підтримки міждисциплінарних досліджень в ССД. Вводяться основні поняття та визначення платформи підтримки міждисциплінарних досліджень в ССД та її системних характеристик, досліджуються проблеми створення платформи та її ефективного використання. Результати аналізу Світової системи даних як об'єкту управління дозволили сформулювати вимоги до показників доступності, оперативності, надійності платформи, обґрунтувати перспективність її створення. Показано, що проблема створення платформи декомпонується на підпроблеми інтеграції застосунків і джерел даних, узгодження даних, впровадження проблемно-орієнтованого апарату інтелектуального аналізу даних та швидкого створення інформаційних ресурсів.

Наведено результати огляду існуючих рішень, встановлено, що жодне з них не реалізує всіх функцій платформи підтримки міждисциплінарних досліджень ССД, обґрунтовано перспективність створення вітчизняної платформи і відібрано моделі, методи, інструменти і технології, які можна використати з цією метою.

Поставлена задача створення платформи підтримки міждисциплінарних досліджень в ССД і уточнені завдання дослідження, спрямовані на побудову відповідного комплексу моделей і методів інтеграції джерел даних і застосунків, генерування проблемно-орієнтованих застосунків для аналітичного оброблення міждисциплінарних даних, обґрунтування архітектури платформи, визначення структури і принципів функціонування та взаємодії її окремих компонентів і особливостей використання на прикладі побудови стратегій та сценаріїв сталого розвитку України та прогнозування пандемій.

Другий розділ присвячено розробленню концепції платформи та моделей і методів інтеграції застосунків. Доступ з дружнім інтерфейсом, який не вимагає спеціальних знань в галузі ІТ, до успадкованої сукупності джерел даних і застосунків вимагає сучасного ІТ-середовища, яке інтегрує виокремлені системи збору, збереження і оброблення даних ССД на *технологічному* (єдиний стек протоколів, форматів передачі даних), *змістовному* (єдиний формат описання семантичної структури джерел) і *функціональному* (використання різномірних сервісів в єдиній структурі для розв'язання поставлених користувачем задач) рівнях.

В основу розподіленої системи покладено відомий спосіб взаємодії сервісів – оркестрацію, який дозволяє групувати базові сервіси в ієрархічно скомплексовані системи, підпорядковуючи їх адміністраторам – сервісам «оркестраторам» і об'єднуючи сервіси за потрібними для ССД ознаками (галузь науки, функціональність, регіональне розташування, тощо). Сервіси-оркестратори

отримують повноваження відповідно до політик міжнародного обміну даними. Простий запит виконується одним сервісом, а складний запит, притаманний міждисциплінарним дослідженням, – композицією сервісів. Для пошуку сервісів і їх композиції кожен сервіс-оркестратор має інформацію щодо функціональності підпорядкованих сервісів і крім того, вони можуть координувати свої можливості в процесі планування і реалізації запитів користувачів. Для включення нового сервісу вноситься у реєстр його опис і він закріплюється за певним сервісом-оркестратором.

Користувачі, які знають інформаційні потреби, можуть скласти відповідні ланцюжки запитів до оркестраторів. Для інших користувачів система забезпечує: підтримку формування запитів в термінах певної предметної області; планування їх виконання; організацію взаємодії базових сервісів в системі. Для цього в ядро системи включені інтелектуальні агенти, які реалізують логіку її функціонування, що базується на методах логічного виведення. Запит користувача виконується після формування виводу, який містить вказівники на потрібні джерела даних і сервіси. Деревовидна структура виводу дозволяє виконувати досить складні композиції сервісів. Потрібні для цього залежності описуються в аксіомах, які вводяться в базу знань при реєстрації відповідних сервісів в системі. Схема взаємодії компонентів запропонованого рішення наведена на рис. 1.

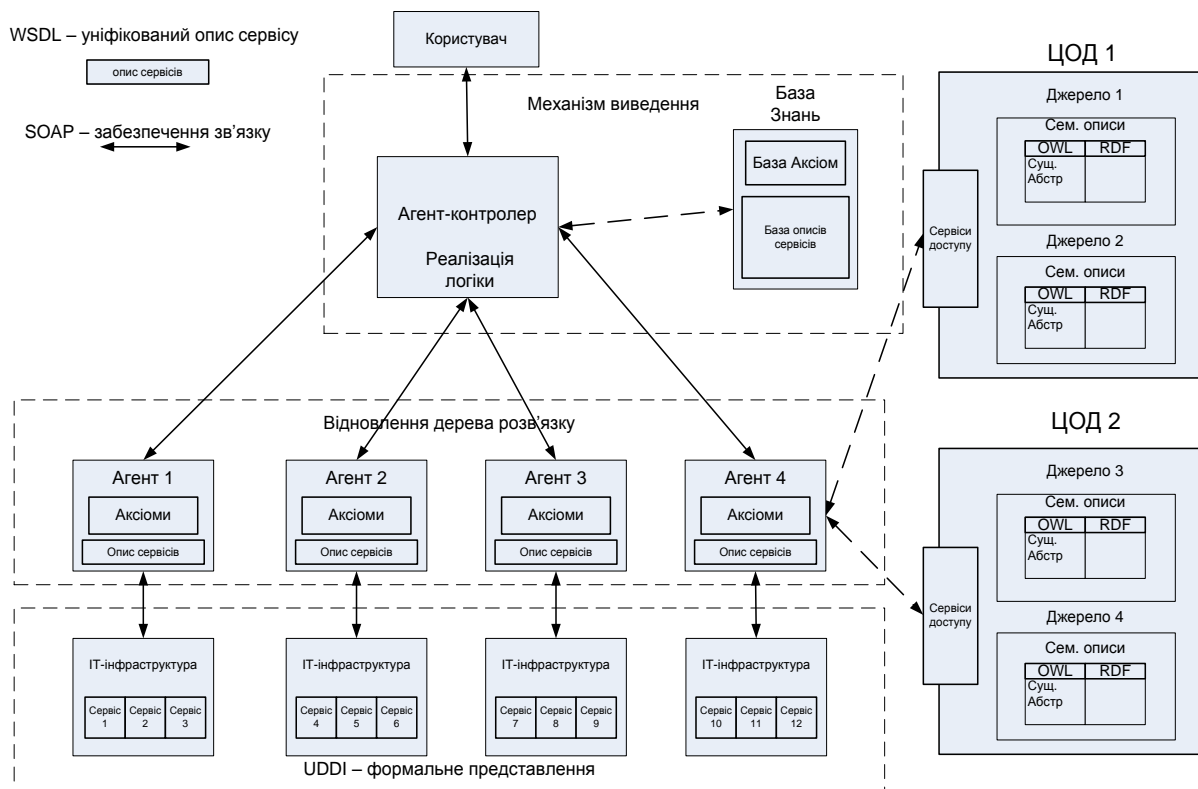


Рисунок 1 – Схема взаємодії компонентів інтеграції застосунків

У рішенні, яке пропонується, UDDI використовується для створення реєстру сервісів, WSDL – для уніфікованого їх описання, SOAP – для обміну повідомленнями, BPEL – для оркестрації сервісів. Опис на WSDL кожного зареєстрованого в системі сервісу нижнього рівня і пов'язаних з ним аксіом заноситься в базу знань системи і безпосередньо в базу знань Агента-Контролера (вибирається згідно з політиками ССД, наприклад за критерієм географічного розташування чи мінімізації обміну

даними), якому новий сервіс підпорядкований.

Агент-оркестратор: отримує запит користувача; опитує агентів (з метою пошуку потрібних сервісів і ресурсів); будує вивід (враховуючи описані аксіомами функції сервісів і правила виведення); передає вивід на нижній рівень для реалізації.

Агент рівня сервісів: виконує виклики необхідних сервісів із свого набору; направляє оброблені дані у відповідності з ланцюжком дій (згідно дерева вирішення задачі, відновленого механізмом відновлення дерева вирішення на основі виводу).

Для реалізації такої взаємодії сервісів у ССД в роботі використовується логічний підхід. Необхідні для його реалізації формальна мова, аксіоми і правила виведення, метод виведення, алгоритм роботи механізму побудови дерева виконання запиту користувача на основі виводу наведені саме у розділі 2 роботи.

Для інтеграції джерел даних і створення складних запитів, що передбачають використання даних з різних джерел, існуючих застосунків і реалізації додаткових алгоритмів оброблення даних у роботі використовується рішення класу ЕІ з посередником і web-орієнтованою архітектурою, загальна схема якого наведена на рис. 2. Використання дескриптивних логік для описання джерел даних, модульність, готовність до приєднання ланцюжків сервісів для додаткового оброблення даних приводять до розподіленої системи з центральним вузлом, який працює на сервері

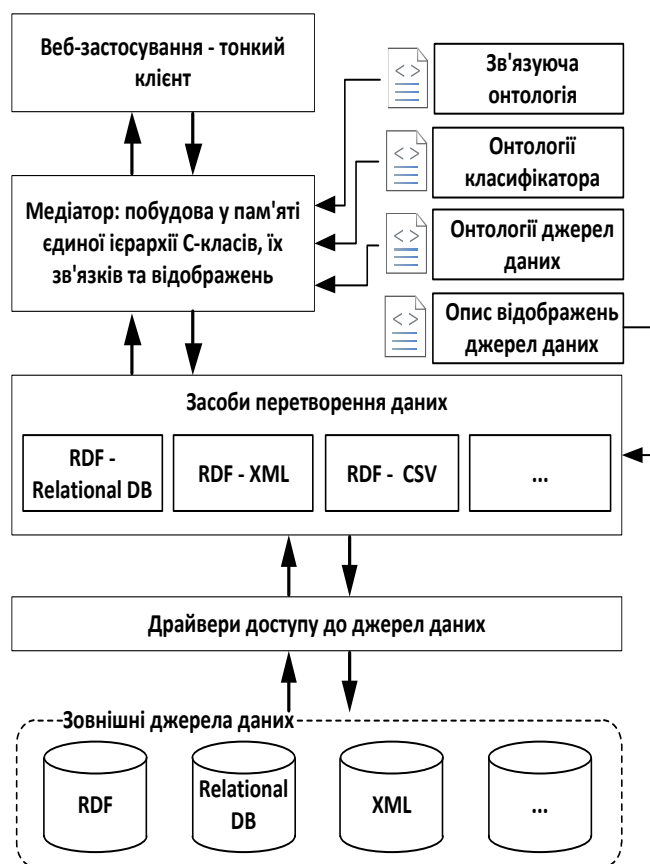


Рисунок 2 – схема інтеграції джерел.

Необхідні для реалізації рішення моделі і методи розглянуті у розділах 2 і 3.

У заключній частині розділу 2 розглянуті складові реалізації логічного підходу до інтеграції застосунків - формальна логічна система, метод виведення, алгоритм побудови дерева виконання запиту користувача на основі виводу.

Формальна система платформи підтримка міждисциплінарних досліджень має забезпечувати визначення потрібних користувачам даних з інших галузей науки. Використовуючи онтології, знання і дані із різних галузей класифікуємо на рівні абстракцій об'єкт – тип (сутність). Компоненти платформи використовують універсуми об'єктів і типів системи, щонайменше опрацьовуючи складові тріади елемент – властивість – відношення, відповідаючи на запити щодо властивостей об'єктів (типів), будуючи об'єкти із заданими властивостями, оцінюючи стан об'єктів і наслідки певних змін в об'єктах і відношеннях універсумів.

Відповідно на кожному з універсамів повинні бути визначені системи відношень і операцій на об'єктах і типах, а, крім того, відношень між елементами цих універсумів, насамперед належності об'єктів типам. Відношення на універсамі типів носять узагальнюючий характер і використовуються для підвищення ефективності роботи на універсамі об'єктів. Система операцій, насамперед перетворення типів, утворення об'єктів типів з заданими властивостями, розпізнавання об'єктів типів, оцінювання властивостей об'єктів з урахуванням відношень, дозволяє визначити в рамках платформи певні процеси за рахунок впровадження таких типів як ситуація, подія, дія, проблема.

У роботі реалізація логічного підходу до інтеграції застосунків побудована на двох формальних логічних системах - логіках об'єктів (*LO*-система) і типів (*LT*-система). За основу *LO*-системи в роботі прийнята відома *PR*-система - клаузальна логіка першого порядку. Мова цієї формальної системи доповнена системою типів, орієнтованих на організацію взаємодії застосунків для міждисциплінарних досліджень в системі світових центрів даних.

До об'єктного типу додано *class*, *property*, *dependency*. З урахуванням очевидних уточнень, пов'язаних із доданими типами, визначення термів типу, формул для індивідів, специфікаторів, передумов і постумов, специфікаторів методів і проблем залишаються у силі. Відповідно клаузи, вирази вигляду $\Pi \rightarrow \Lambda$, де Λ і Π становлять відповідно одну і послідовність атомарних формул, поділимо на інтегральні (для об'єктів і типів), об'єктні (для об'єктів), типізовані (для типів).

Базу знань цієї системи складають онтологія аксіом, які зображують методи сервісів системи, і онтології джерел даних, описаних на мові OWL на базі RDF. Також до бази знань належать правила виведення, необхідні для отримання потрібного результату у випадках, коли необхідно комбінувати методи як одного, так і різних сервісів:

1. Якщо $d_1: \langle \tau_1, \tau_2 \rangle$ і $d_2: \langle \tau_3, \tau_1 \rangle$ то $d_1: \langle d_2, \tau_2 \rangle$
2. Якщо $d_1: \langle \tau_1, \tau_2 \rangle$ і $d_2: \langle \tau_3 \wedge \tau_4, \tau_1 \rangle$ то $d_1: \langle d_2, \tau_2 \rangle$
3. Якщо $d_1: \langle \tau_1 \wedge \tau_2, \tau_3 \rangle$ і $d_2: \langle \tau_4, \tau_1 \rangle$ і $d_3: \langle \tau_5, \tau_2 \rangle$ то $d_1: \langle d_2 \wedge d_3, \tau_3 \rangle$
4. Якщо $d_1: \langle \tau_1, \tau_2 \rangle$ і $d_2: \langle \tau_3 \vee \tau_4, \tau_1 \rangle$ то $d_1: \langle d_2, \tau_2 \rangle$
5. Якщо $d_1: \langle \tau_2 \wedge \tau_2, \tau_1 \rangle$ то $d_1: \langle \tau_2, \tau_1 \rangle$
6. Якщо $d_1: \langle \tau_1, \tau_2, \tau_3 \rangle$ і $d_2: \langle \tau_4, \tau_1 \rangle$ і $d_3: \langle \tau_5, \tau_2 \rangle$ то $d_1: \langle d_2, d_3, \tau_3 \rangle$
7. Якщо $d_1: \langle \tau_1, \tau_2, \tau_3 \rangle$ і $d_2: \langle \tau_2, \tau_4 \rangle$ і $d_3: \langle \tau_3, \tau_5 \rangle$ то $d_2: \langle d_1, \tau_4 \rangle$ і $d_3: \langle d_1, \tau_5 \rangle$

За основу *LT*-системи в роботі прийнята відома *TS*-система - клаузальна логіка вищого (скінченного) порядку. Мова цієї формальної системи є розширенням мови *LO*-системи, доповненим константами і змінними, орієнтованими на організацію взаємодії застосунків для міждисциплінарних досліджень в ССД.

До схем додано можливість отримання їх на основі об'єктів введених вище нових типів *class*, *property*, *dependency*. З урахуванням очевидних уточнень, пов'язаних із доданими типами, визначення термів, схем, специфікаторів схем, формул, клауз залишаються у силі.

Базу знань системи складають схемна компонента (множина специфікаторів схем), схемно-проблемна компонента (множина специфікаторів схем розв'язання проблем), схемна трансформаційна компонента (множина правил побудови схем розв'язання проблем), компонента трансформацій у системі *class*, *property*, *dependency* (множина правил побудови схем розв'язання проблем на основі залежностей онтологій).

В дисертаційній роботі отримав подальший розвиток метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в універсумі типів з наступною деталізацією виводу в системі універсумі об'єктів. Запропонований метод базується на типізації тверджень і аналогії. Спочатку будується вивід у *LT*-системі для опрацювання запиту на універсумі об'єктів. Тут спочатку використовується абстракція для отримання опису проблеми у термінах універсуму типів. А потім будується вивід у *LO*-системі для опрацювання запиту на універсумі об'єктів. При цьому вивід у *LT*-системі використовується для управління процесом виведення *LO*-системі. Як і при будь-якій типізації тверджень підвищується ефективність виведення, оскільки відсікається більша частина безперспективних гілок виводу як на універсумі об'єктів, так і на універсумі типів. Оскільки запропонована типізація будується природним чином на основі семантичних описів джерел даних, зменшуються витрати на формування і підтримку бази знань системи.

Упорядковані лінійні виводи мають бажані властивості, використання яких дозволило побудувати механізми виведення для агентів платформи, які задовольняють вимоги до неї. Після формулювання користувачем запиту, система буде вивід – по суті, перевіряючи можливість задоволення запиту за рахунок наявних ресурсів. Це завдання виконують агенти-оркестратори як важлива складова інтелектуального ядра системи. Маючи доступ до переліку всіх сервісів, їх описів і аксіом, які й специфікують можливості застосування сервісів, вони запускають процес формування виводу. Тоді й відбувається побудова ланцюжка сервісів, здатного забезпечити потрібний користувачу результат, розпочинаючи з отримання даних із відповідних їх джерел.

Першу із множини *T* трійку упорядкованого лінійного вивід формуємо із постулату сформульованої проблеми, сумісної аксіом з бази знань, результату застосування правила виведення. Третя вершина першої трійки стає першою вершиною другої трійки і так доти, поки не буде досягнуто термінальної вершини – передумови, вказаної в проблемі. Якщо під час пошуку аксіом з бази знань для чергової вершини сумісних аксіом знаходиться декілька, то кожна з цих аксіом використовується для подальшої побудови своєї «паралельної» версії виводу. Хоч запропонований метод відсікає частину виводів, після завершення роботи з множини отриманих виводів вибирається вивід з найменшою кількістю трійок вершин і найменшою кількістю застосованих аксіом.

Для отримання з виводу функціональної послідовності дій, яку і буде виконувати платформа, з врахуванням особливостей і характеру сервісів, необхідно задіяти механізм відновлення схеми вирішення із отриманого виводу.

Схема вирішення становить зв'язний орієнтований граф без орієнтованих циклів з паралельними орієнтованими шляхами з кореня у вершини, який має три типи вершин, і задається трійкою $G=\langle V, E, \Theta \rangle$, де $V=V_1 \cup V_2 \cup V_3$, V_1 - множина вершин-методів, V_2 - множина вершин-передумов і вершин-постумов, V_3 - множина вершин даних, в яких відбувається об'єднання або розщеплення даних; E - множина ребер; Θ - підмножина декартового добутку $E \times V \times V$, яка визначає співвідношення ребер і пар вершин. Схема визначає послідовність і взаємозв'язок по даним дій, які будуть виконані платформою для отримання потрібного користувачу результату.

Алгоритм відновлення схеми $G=\langle V, E, \Theta \rangle$ вирішення на основі виводу $\text{Result} = \langle V, T \rangle$.

Крок 1. Визначення термінальної трійки – такої трійки, яка містить вершину k , що не є ані першою, ані другою складовою жодної з трійок множини T .

Крок 2. Визначення методу за однозначною відповідністю формулам вершин постумови та передумови аксіоми методу.

Крок 3. Вершини, ребра та їх співвідношення заносяться до G .

Крок 4. Вилучивши оброблену трійку з множини трійок вершин T , виконуємо крок 2. Якщо множина пуста – відновлення дерева вирішення завершено.

Після завершення відновлення дерева вирішення, воно подається на вхід виконавчої компоненти. При цьому можливе паралельне виконання методів.

Третій розділ присвячено моделям і методам інтеграції джерел даних, міжагентної взаємодії та системного узгодження даних різної природи. Для *інтеграції джерел даних* в роботі розроблені механізми федерування, побудови запитів, об'єднання результатів як основі організації, планування і здійснення підтримки міждисциплінарних досліджень в Світовій системі даних. Саме завдяки цим моделям і методам програмуєний користувач може створювати ефективні схеми виконання складних запитів до системи, що передбачають використання багатьох баз даних, існуючих застосунків і реалізації нових алгоритмів обчислень. По суті мова йде про створення нових застосунків.

Для описання специфікації семантики джерел даних обрані дескриптивні логіки. Як інструмент подання узагальнених специфікацій обрано онтологічні специфікації. Позначимо онтологічний клас і властивість в контексті семантики як C -клас і C -властивість відповідно. Щоб застосувати стек технологій Semantic Web для інтеграції джерел даних використовуємо абстрактну модель даних у вигляді орієнтованих графів «суб'єкт-предикат-об'єкт» RDF, мову описання онтологій OWL і мову запитів до RDF-подібних даних SPARQL.

Для високорівневого описання доступних в системі предметних областей в роботі створена центральна онтологія-класифікатор, яка складається з ієрархії C -класів і не залежить від джерел даних. Вона може бути розширена іншими онтологіями такої ж структури або парами онтологія-відображення. Кожна така пара забезпечує підключення для конкретного джерела даних. Для цього онтологія описує структуру і зв'язки його C -класів і C -властивостей, а відображення вказує співвідношення його C -класів і C -властивостей до реальної структури джерела даних,

а також параметри підключення до нього. Таким чином, запит до джерела даних може формуватися в термінах *C*-класів та *C*-властивостей. Оскільки різні джерела даних можуть мати спільні *C*-властивості, в роботі створена онтологія для встановлення зв'язку між такими різних джерел даних за допомогою стандартних предикатів OWL. На рівні застосунку зв'язування онтологій здійснюється за рахунок унікального для кожного *C*-класу Uniform Resource Identifier (URI).

Склад і взаємодія компонентів системи інтеграції джерел даних наведені на рис. 3. Компоненти функціонують на основі моделей і алгоритмів, розроблених в роботі. Розглянемо послідовність роботи компонентів над запитом.

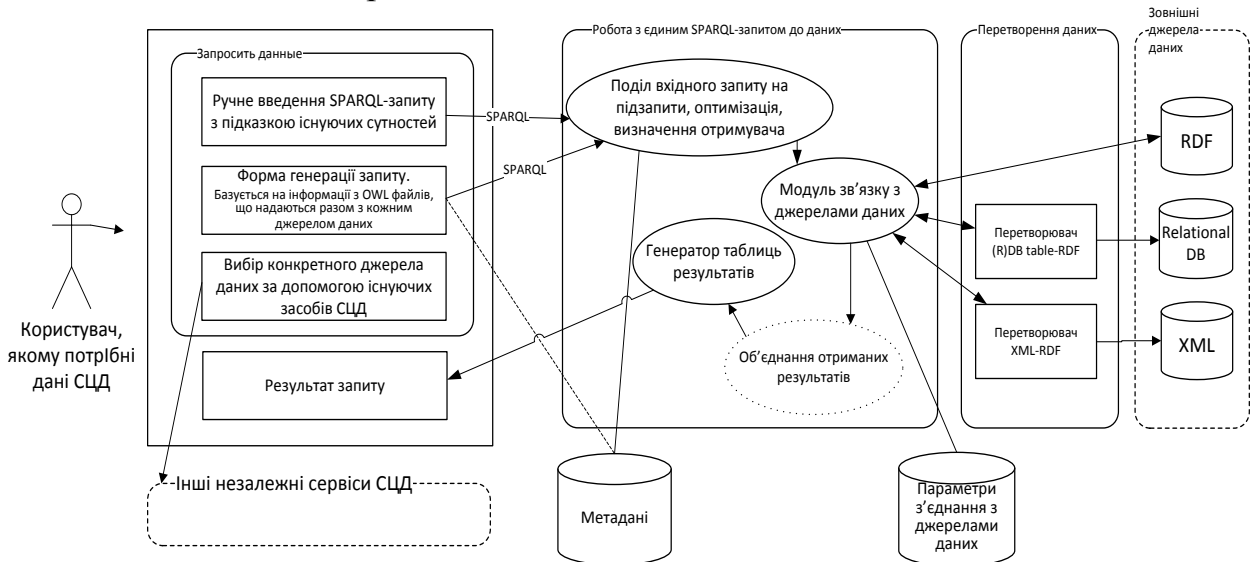


Рисунок 3 – Склад і взаємодія компонентів системи інтеграції джерел даних

Етап 1. Генерування на основі онтологічного опису джерел даних та метаданих SPARQL-запитів за допомогою конструктора або отримання готового запиту від користувача в термінах *C*-класів, *C*-властивостей, їх зв'язків та обмежень в глобальному просторі імен. З урахуванням характеристик джерел даних, правил виведення, що визначають можливі операції над *C*-класами та їх *C*-властивостями, формується план виконання запиту (визначаються джерела даних і способи об'єднання результатів). Після цього запит трансформується в єдиному просторі імен на основі описів відображень між джерелами даних, конвертується в запити до окремих джерел даних у відповідному синтаксисі, наприклад, у форматі SQL.

Розроблено модель для формування користувачем єдиного запиту до множини джерел даних у єдиному просторі імен. Користувач оперує ієрархією *C*-класів, зв'язками між ними (описані owl-предикатами SubclassOf і SameAs), а також *C*-властивостями *C*-класів і їх обмеженнями. На рівні логіки застосунку визначено, які джерела даних пов'язані з якими *C*-класами і *C*-властивостями. Послідовність дій при формуванні запиту: 1) вибір необхідного *C*-класу в дереві класифікатора; 2) встановлення обмежень на значення його *C*-властивостей (за їх наявності) за допомогою констант або логічних виразів; 3) встановлення зв'язків з *C*-властивостями інших *C*-класів з формуванням Join-запиту (опціонально).

Таким чином, на виході може бути отримано один або більше одного *C*-класів, набір *C*-властивостей і обмежень на них, *C*-властивості, через які треба об'єднувати

значення для отримання кінцевого результату. У межах цієї моделі можуть бути оброблені чотири типи стандартних запитів до розподілених джерел даних: *запит до однієї сутності* (SPARQL-запит буде містити URI потрібного *C*-класу, за яким закріплене тільки одне джерело даних); *запит, в якому сутність рознесена в декількох джерелах даних* (*Union*); *запит з об'єднанням сутностей* (*Join*); *запит на узагальнення* – запит на отримання *C*-класів, *C*-властивостей та їх зв'язків, які не підлягають об'єднанню, але, згідно опису в онтології-класифікаторі, вони є підкласами обраного користувачем *C*-класу.

Реалізація наступних кроків вимагає зв'язної структури онтологій і підключень до джерел даних. У роботі запропоновано алгоритм побудови зв'язної ієрархії із онтології-класифікатора та інших онтологій і описів джерел даних:

- А. Завантаження онтології-класифікатора, побудова головного дерева *C*-класів.
- Б. Завантаження і приєднання до головного дерева розширюючих онтологій.
- В. Завантаження онтологій-описів джерел даних.

Г. Приєднання завантажених онтологій-описів джерел даних до головного дерева, співставлення співпадаючих *C*-класів і їх *C*-властивостей та їх специфікація як доступних у кількох джерелах.

Д. Завантаження зв'язуючої онтології, її проекція на головне дерево (результатом є граф, що пов'язує листки, тобто *C*-властивості головного дерева).

Є. Завантаження описів відображень джерел даних і встановлення їх зв'язків з *C*-класами та *C*-властивостями.

Вигляд дерева *C*-класів, *C*-властивостей, їх зв'язків між собою і з джерелами даних, створеного для застосунку, наведено на рис. 4.

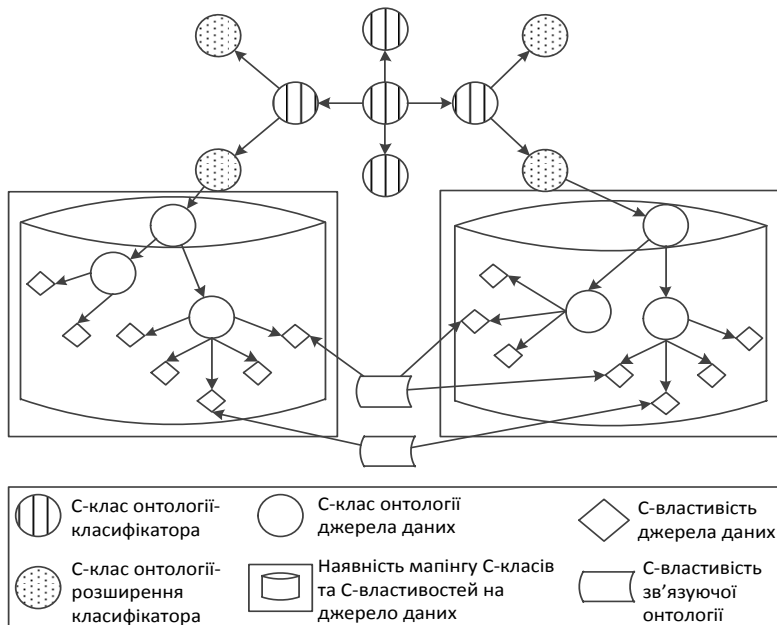


Рисунок 4 – Приклад об'єктної моделі

Етап 2. Парсинг, переписування і оптимізація запиту. Виконується на основі онтологічного опису джерел даних та метаданих. Ефективність системи інтеграції джерел даних визначається компонентом планування виконання запиту та його оптимізації. Оскільки джерела даних функціонують автономно, статистичні дані про їх наповнюваність отримати не завжди можливо. До того ж доцільність виконання надскладних запитів до підключених джерел даних, враховуючи їх специфіку у світовій системі даних і невелику кількість поєднуваних властивостей, є сумнівною.

Тому в роботі запропонований алгоритм планування на основі статистичної інформації, яку надають постачальники даних (коефіцієнтів ємності джерел даних). Обробник генерує дерево-план виконання запиту, яке забезпечує відпрацювання запитів до джерел з найменшими витратами ресурсів, допускаючи паралельне їх виконання, об'єднання результатів. Наведемо кроки цього алгоритму:

Крок 1. Отримання запиту у єдиному просторі імен (джерела даних, C-властивості, коефіцієнти ємності кожного джерела даних);

Крок 2. Знаходження добутку коефіцієнтів ємності кожної пари джерел даних, підзапити до яких підлягають об'єднанню;

Крок 3. Виділення пар джерел, підзапити до яких підлягають об'єднанню;

Крок 4. Впорядкування пар у порядку зростання добутку коефіцієнтів ємності;

Крок 5. Виділення джерел даних, запити до яких можуть бути виконані паралельно і мають найменші коефіцієнти ємності;

Крок 6. Паралельне виконання запитів до виділених на кроці 5 джерел даних;

Крок 7. Виділення з результатів кроку 6 даних, за якими виконується перетин із джерелом-парою та передавання їх як параметра в запиті до цього джерела-пари;

Крок 8. Представлення отриманого перетину результатів запитів до кожної пари джерел даних як віртуального джерела даних, його коефіцієнту ємності;

Крок 9. Виконання кроків 3 – 8 доки не буде отримано єдиний кортеж, який і стане результатом виконання запиту користувача.

Етап 3. Разбиття запиту на підзапити. Формуються запити до джерел даних. Поля таблиць джерела даних приводяться у абстрактну модель даних у вигляді графу «суб'єкт-предикат-об'єкт» і навпаки за допомогою опису відображень (мапінгу). Представимо опис відображення як $f: R \leftrightarrow S$, де R – вихідна структура джерела даних, S – абстрактне семантичне представлення у вигляді графу. Використовується відомий алгоритм безпосередньої конвертації запитів. Запити передаються у джерела за допомогою стандартних адаптерів, наприклад, JDBC.

Етап 4. Виконання підзапитів на вибраних джерелах даних. Джерела даних автономно отримують і виконують запити, повертаючи набори записів як результат.

Етап 5. Очікування отримання відповідей від усіх джерел даних.

Етап 6. Опціональне об'єднання результатів за допомогою Inner Join. Отримані від джерел даних результати виконання SQL-або SPARQL-запитів опрацьовуються згідно вище сформованого плану виконання запиту шляхом виконання операцій об'єднання, перетину, передавання сервісам-обробникам і т.д. Запропоновано також рішення для взаємодії з джерелами даних, які надають XML-файли.

Етап 7. Повернення користувачу результат виконання запиту - таблиці або множини таблиць (якщо об'єднання неможливе або не потрібне користувачеві).

Запропоноване рішення не вимагає від користувача знання структури джерел, характеристик і особливостей мов запитів. За критеріями гетерогенність, розподіленість і автономність воно переважає рішення побудовані на множинних (multidatabase), федеративних (federated database) і монолітних базах даних.

Моделі і методи *міжагентної взаємодії* залежать від подання мультиагентної системи та структурної організації її агентів. Прийнятий підхід оснований на застосуванні ідеї колективної діяльності сукупності агентів – мультиагентних систем (МАС). Фундаментальними ознаками МАС приймаємо її здатність виконувати у певному середовищі автономні дії з метою досягнення поставлених під час проектування цілей і включення набору автономних агентів, що можуть взаємодіяти між собою. Необхідність оцінювання поведінки та продуктивності, планування і перепланування дій згідно визначених цілей вимагає прийняти до уваги такі

загальновизнані властивості MAC, як автономність, соціальність, реактивність та проактивність, оптимальність, адаптивність.

Для визначення семантики даних використовується Web Ontology Language (OWL). Мультиагентний підхід забезпечує придатність для подальшого розширення функціональності шляхом впровадження агентів, що забезпечують певні перетворення даних, інтеграцію даних відповідно до їх семантично однорідних фрагментів і візуалізацію у вигляді карт, графіків і інших уявлень.

Для реалізації MAC обране середовище JADE (Java Agent DEvelopment Framework). Розроблена в дисертаційній роботі архітектура передбачає існування $n \geq 1$ агентів, які обслуговують джерело даних і мають унікальні для їх предметної області OWL-онтології і файли XML, що визначають параметри підключення до джерела, правила проектування словника на джерело даних і параметри функціонування агента в інтегрованій системі.

Агенти обмінюються інформацією через агента - «дошку оголошень», користувач повідомляє про наявність нових запитів або отримує результати через web-інтерфейс за допомогою агента, що організує зв'язок web-сервера з агентним середовищем. Агенти можуть звертатися до джерел як локально так і віддалено. Міграція агентів уможливорює оптимізацію логічного виведення над словником даних при обробленні запиту користувача і розподіл навантаження. Пошук агентів MAC і надаваних сервісів здійснюється на підставі каталогу DF (Directory Facilitator). Розроблені агенти для перетворення отриманих результатів на запит користувача і запиту на підмножині природної мови в запит SPARQL. Інтероперабельність даних і можливість виділення семантично однорідної інформації забезпечуються парадигмою Linked Data. Для зіставлення класів і властивостей, визначених у словнику OWL Tbox (terminological component) і їх екземплярів в джерелі даних Abox (assertion component) використаний D2RQ. Зв'язок з джерелом даних здійснюється за допомогою jdbc драйверів.

На *системному рівні* організації MAC складовими частинами є множина агентів та їхнє середовище. Кожен із агентів здатен аналізувати стан середовища, у якому знаходиться, та генерувати певний вплив на це середовище. Вхідні та вихідні дані агента A мають свої алфавіти (X та Y відповідно), а сам агент може розглядатися як перетворення над даними в цих алфавітах $A: X \rightarrow Y$. Тут $X = X_D \times X_C$, де алфавіт X_D використовується для опису вхідних аргументів перетворення, алфавітом X_C - директив контролю перетворення. Якщо агент не надає відомостей про свій вихідний алфавіт, то він здатен лише отримувати вхідні дані. Для таких агентів вихідний алфавіт рівний $O = \{\emptyset\}$. Можливі три типи агентів:

- 1) агенти-перетворення (з повними вхідним та вихідним алфавітом), $A: X \rightarrow Y$;
- 2) агенти-споживачі (з пустим алфавітом вихідних даних), $A: X \rightarrow O$;
- 3) агенти-генератори (з пустим алфавітом вхідних аргументів), $A: O \times X_C \rightarrow Y$.

Для виконання перетворення $X' \rightarrow Y'$ над аргументами x агенти отримують дані, представлені кортежем $R = \langle X', Y', x \rangle$. Адаптивність та автономність системи вимагає самооцінювання власного стану та винесення рішення про необхідність змін у роботі. Стан агента описується вектором $\bar{s} \in R^p$, який містить значення кожного з p параметрів роботи агента.

Структурний рівень організації агента агента A визначається набором елементів $a_i, i = \overline{1, n}$, кожен з яких може виконувати перетворення f_{a_i} над підмножиною вхідного алфавіту агента X^i . При цьому результатом такого перетворення є підмножина вихідного алфавіту агента Y^i . Тобто, $A = \{a_i | f_{a_i}: X^i \rightarrow Y^i\}, i = \overline{1, n}, X = \bigcup_i X^i, Y = \bigcup_i Y^i$. Семантика такого перетворення визначається парою вхідного та вихідного алфавітів та символом перетворення (X^i, Y^i, f_{a_i}) . Дані про елементи заносяться до внутрішнього реєстру агента. Набір елементів в агенті може змінюватися завдяки реконфігурації.

Внутрішня структура агента представлена орієнтованим графом $G = (V, E)$, а множина перетворень елементів агента - підмножиною прямого добутку його вхідного та вихідного алфавітів $A \subseteq X \times Y = \bigcup_i X^i \times \bigcup_i Y^i$. Тут $V = A$, а $E = \{(j, k): Y^j \subseteq X^k\}, j = \overline{1, n}; k = \overline{1, n}$.

При появі заявки $R = \langle X', Y', x \rangle$ агент виконує відповідне перетворення за умови наявності елементів, вхідний та вихідний алфавіти яких відповідають заявці, що надійшла

$$\begin{cases} X' \subseteq X, \\ Y' \subseteq Y; \end{cases} \Rightarrow \exists \{j | j \in [1, n]\}, \{k | k \in [1, n]\}: X' = \bigcup_{\{j\}} X^j \cap Y' = \bigcup_{\{k\}} Y^k.$$

Достатньою умовою виконання заявки агентом є існування не пустої множини шляхів $\{P_{j,k}\}$ між відповідними вершинами j та k : $P_{j,k} = \{p_{j,k}\} = \{a_j, \dots, a_k\}$.

Оскільки вважається, що передача даних по дузі графа G пов'язана зі зміною стану агента $(\Delta \bar{s})$, існує множина функцій $\{\varphi: E \rightarrow R^p\}$, які обчислюють вартість кожного з переходів у графі для агента. Таким чином, структура S агента описується множиною його елементів-перетворень A , вектором стану \bar{s} та множиною функцій оцінки вартість кожного з переходів у графі G : $S = \langle A, \bar{s}, \{\varphi\} \rangle$.

У випадку існування у графі G декілька маршрутів між парами вершин j та k , що дає можливість зміни шляху обробки заявки в залежності від вхідних даних x , поточного набору елементів $A = \{a_i\}$, та значень параметрів стану агента \bar{s} , у роботі у роботі розроблені засоби розв'язання відповідної задачі оптимізації.

Процес оброблення заявки R в агенті описується як просування пакету C^z , який у момент часу τ становить четвірку $C^\tau = \langle u, M^\tau, d^\tau, P^\tau \rangle, P^\tau \subseteq A$, де $u \in N$ є унікальним ідентифікатором поточного запиту, M - алфавіт даних, які обробляються, d - поточні дані для оброблення елементом агента, P - шлях, який даний пакет пройшов в агенті. При $\tau = 0$ маємо $M^\tau = X', d^\tau = x, P^\tau = \{\}$.

Після обробки пакета елементом a_t ($t \in [1, n]$) агента, частини M та d змінюються відповідно до вихідного алфавіту поточного елемента та результатів здійсненого перетворення: $M^{\tau+1} = Y^t, d^{\tau+1} = f_{a_t}(d^\tau), P^{\tau+1} = P^\tau \cup a_t$.

При подальшій передачі пакета $C^{\tau+1}$ наступним елементам агент визначає адресата, який має відповідний вхідний алфавіт. Для прийняття рішень агент формує вибірку $\{a_i | M^\tau \subseteq X^i\} \subseteq A, i = \overline{1, m}$ із m доступних елементів, керуючись результатами аналізу поточного стану системи.

Ситуація з декількома маршрутами можлива, якщо вхідні алфавіти декількох елементів агента перетинаються: $m > 1 \Rightarrow \exists j, k \in [1, n]: X^j \cap X^k \neq \{ \}$. Зміни у роботі системи здійснюються шляхом модифікації рішень стосовно доставки пакета через граф G агента. Для просування пакету агент вибирає маршрут з найменшою вартістю. Коли такий маршрут невідомий, виконується широкомовна передача пакета елементам, які можуть обробити дані, що передаються. Із усіх можливих маршрутів, утворених таким чином, вибирається той, який забезпечує мінімальний час проходження пакета. Такі маршрути запам'ятовуються і використовуються в подальшому для оброблення подібних заявок. На підставі поточного стану і набору функцій φ , агент може прийняти рішення про «небажаність» проходження пакету певною дугою графа G . У такому випадку він корегує вибраний маршрут.

Отже, агент може змінити спосіб обробки вхідних заявок у залежності від власних цілей, які визначаються функцією φ . Така реконфігурація маршруту здійснюється у випадку появи в структурі агента елемента, який входить до альтернативного маршруту, або зміни вартості маршрутів обробки заявки.

Розроблені в роботі моделі і методи *системного узгодження даних різної природи* призначені для приведення даних до єдиної семантики, єдиного діапазону значень і єдиних одиниць виміру за умови мінімізації інформаційних втрат такого узгодження. Потреба в них викликана тим, що використовувані в міждисциплінарних дослідженнях дані мають різну природу, визначену їх об'єктивним змістом, цільовим призначенням і способом отримання. Узгодженню підлягають кількісні і (або) якісні оцінки властивостей деякої сукупності об'єктів $O = \{o_i\}, i = \overline{1, n}$, де o_i – значення ідентифікаторів номінальної шкали об'єктів представленої сукупності, як результати відображень вигляду:

$$O \xrightarrow{I_j} X^j, j = \overline{1, m} \quad (1)$$

де $I_j, j = \overline{1, m}$ – відображення, визначені на множині об'єктів ($D(I_j) = O$). Їх області значень – це області визначення показників X^j , тобто $E(I_j) = D(X^j)$.

Для узгодження даних в роботі запропоновані оцінки інформаційних втрат і зростання невизначеності моделі, що супроводжують цей процес, методологія кількісного оцінювання узгодженості даних різної природи і методи узгодження.

Інформаційні втрати оцінюються при перетворенні вимірювальних шкал. Вводяться емпірична $E = (O, R)$, де $R = \{r_k\}, k = \overline{1, m}$ – множина відношень на множині O , і вимірювальна системи $M = (O', R')$, де $O' = \{o'_k\}, k = \overline{1, n'}$ – множина символів, $R' = \{r'_l\}, l = \overline{1, m'}$ – множина допустимих відношень на них, відповідність між якими задається сюр'єкцією $g: E \rightarrow M$. Вимірювальний експеримент полягає у визначенні стану першої за допомогою другої.

Шкала вимірювань становить трійку $S = (E, M, g)$, а оцінювання інформаційних втрат при вимірах у цій шкалі пов'язане з невизначеністю зворотного відображення $g^{-1}: M \rightarrow E$. Інформація, отримана в результаті безпосереднього з'ясування стану емпіричної системи, визначається як $I_E = H(E)$, де $H(E)$ – її власна ентропія. Але, як правило, стан емпіричної системи можна визначити лише опосередковано з

допомогою шкали $S = (E, M, g)$. В цьому випадку емпірична і вимірювальна системи розглядаються як залежні, і має місце співвідношення:

$$I_{M \rightarrow E} = H(E) - H(E|M) \quad (2)$$

де $H(E|M)$ – умовна ентропія, що характеризує невизначеність стану емпіричної системи у випадку, а стан вимірювальної системи визначений повно.

Таким чином, інформаційні втрати можна оцінити за допомогою величини.

$$\Delta I(M, E) = I_E - I_{M \rightarrow E} = H(E|I). \quad (3)$$

Процес перетворення даних із шкали S_1 в шкалу S_2 розглядаємо як процес вимірювання, в якому S_1 відіграє роль емпіричної системи, а S_2 – вимірювальної, тобто можна визначити нову шкалу $T_{S_1 \rightarrow S_2} = (S_1, S_2, \varphi)$, де $\varphi: S_1 \rightarrow S_2$.

Шкали S_1 і S_2 еквівалентні, якщо для $\exists \varphi \in \Phi: S_1 \rightarrow S_2$ має місце $\Delta I(S_1, S_2) = 0$. Відносно класів $\Phi = \{\varphi\}$ множина шкал $T_{S_1 \rightarrow S_2} = (S_1, S_2, \varphi)$ поділяється на класи еквівалентності (типи) S_Φ . Перетворення даних в межах шкал одного типу не призводить до інформаційних втрат.

У роботі розглянуто найбільш поширені типи шкал, які дозволяють давати як кількісну, так і якісну оцінку властивостей досліджуваних об'єктів (тут a, b – параметри масштабу і зсуву, $O = X^i \subseteq R, O' = X^j \subseteq R$):

- кількісні ($\Phi = \{\varphi: x_i = ax_j + b\}$, $S_i, S_j \in S_\Phi$, $x_i \in X^i, x_j \in X^j$);
- порядкові ($\Phi = \{\varphi: \forall x_i \leq x_j \Rightarrow \varphi(x_i) \leq \varphi(x_j)\}$, $S_i, S_j \in S_\Phi$, $x_i, x_j \in X^i, \varphi(x_i), \varphi(x_j) \in X^j$);
- номінальні ($\Phi = \{\varphi: \forall x_i \neq x_j \Rightarrow \varphi(x_i) \neq \varphi(x_j)\}$, $S_i, S_j \in S_\Phi$, $x_i, x_j \in X^i, \varphi(x_i), \varphi(x_j) \in X^j$).

Рішення задачі узгодження даних полягає у побудові процедури перетворення $\varphi: S_1 \rightarrow S_2$. Якщо шкали S_1 і S_2 різних типів і $\Delta I(S_1, S_2) = H(S_1) - H(S_2) > 0$, то шкала S_1 «сильніша», ніж шкала S_2 , і при переході $S_1 \rightarrow S_2$ виникають інформаційні втрати. При зворотному ж переході, коли $\Delta I(S_2, S_1) < 0$, в модель вноситься невизначеність $-\Delta I(S_2, S_1)$. Найслабшими є номінальні шкали, найсильнішими – кількісні шкали, порядкові шкали займають проміжне положення.

Оцінки інформаційних втрат при переході від кількісної шкали до порядкової або номінальної формулюється в термінах кількості відліків n вихідної (кількісної) шкали і кількості відліків m цільової (порядкової або номінальної) шкали. Для порядкової і номінальної шкали використовуємо вирази відповідно:

$$\Delta I(n, m) = \sum_{l=1}^m \frac{R(n, l) C_m^l \log(C_{\max(n, l)}^{\min(l, n)})}{m^m}, \quad \Delta I(n, m) = \sum_{l=1}^m \frac{S(n, l) A_m^l \log(A_{\max(n, l)}^{\min(l, n)})}{m^m}. \quad (4)$$

Тут C_m^l , $C_{\max(n, l)}^{\min(l, n)}$ означають числа сполучень, A_m^l і $A_{\max(n, l)}^{\min(l, n)}$ – розташувань,

$$R(n, l) = \sum_{j=0}^l (-1)^j C_l^j (l-j)^n - \text{Моргана, а } S(n, l) = \frac{R(n, l)}{l!} - \text{Стірлінга другого роду.}$$

Перед розробленням методу перетворення даних з мінімальними втратами з метою їх узгодження у роботі виконане дослідження зростання невизначеності моделі

при переході від одних типів шкал до інших. Показано, що невизначеність моделі при переході від «слабших» шкал до «сильніших» може бути зменшена за рахунок використання додаткової інформації про досліджувані об'єкти, наприклад середніх значень будь-якого показника x^j для кожного кластера C^j . Корисність додаткової інформації при перетворенні даних можна оцінити за допомогою виразу

$$U(I, S_1, S_2) = \Delta I(S_1, S_2) - \Delta I((S_1, S_2) | I), \quad (5)$$

де $U(I, S_1, S_2)$ – корисність додаткової інформації I при переході від шкали S_1 до шкали S_2 , $\Delta I(S_1, S_2)$ і $\Delta I((S_1, S_2) | I)$ – втрати інформації при перетворенні даних відповідно без використання і з використанням додаткової інформації I .

Виконані експериментальні дослідження дозволили отримати залежність між довжиною протоколу вимірювань k (кількість об'єктів, що підлягають аналізу), і нормованими значеннями $\Delta I(S_1, S_2)$, $\Delta I((S_1, S_2) | I)$ і $U(I, S_1, S_2)$. Експерименти показали, що саме для критичних довжин протоколів, коли інформаційні втрати є найбільшими, корисність додаткової інформації виявляється найбільшою і складає в відносних одиницях близько 60% від максимального значення інформаційних втрат.

Також перед розробленням методу перетворення даних з мінімальними втратами у роботі розроблено підхід до побудова метрик для оцінки узгодженості даних. Дані представлені у вигляді таблиці «об'єкт-властивість» $X = (x_{i,j})_{i=1, j=1}^{n,m}$ розміром $n \times m$, в якій рядок X_i відповідає набору значень, що характеризує властивості об'єкта o_i , а стовпець X^j задає значення j -ого показника для всієї вибірки об'єктів.

Вибірки об'єктів, часові ряди, проекції властивостей і т.п. відповідають різним підмножинам $V \subseteq X$, які можуть бути задані як композиції на множині елементарних вибірок $V_i \equiv X_i$ – горизонтальних перетинів цієї таблиці, і елементарних проекцій $V^j \equiv X^j$ – вертикальних перетинів тієї ж таблиці.

У роботі проаналізовані кілька моделей, на яких вводяться метрики узгодження. Найчастіше кількісна оцінки узгодженості на основі побудови метричних просторів даних ґрунтується на використанні поділів. У загальному випадку на X можна задати сигма-алгебру S і визначити метричний простір (X, S, μ) , ізоморфний ймовірнісному простору. Тобто незалежно від типів шкал представлення даних відбувається відображення їх значень в номінальні події, визначені на \tilde{X} . Якщо дані представлені в кількісних шкалах, $\tilde{X}_i \in \tilde{X}$ можна розглядати як елементи лінійного простору $E(X^1) \times \dots \times E(X^m)$ або використовувати інші моделі, наприклад кватерніони, на яких вводяться метрики узгодження.

У тому випадку, коли вдається визначити простір з нормою:

$$\|\tilde{X}_i\| = \left(\sum_{j=1}^{|\tilde{X}|} w_j (x_{i,j})^p \right)^{\frac{1}{p}}, \quad (6)$$

де w_j – вагові коефіцієнти об'єктів або показників, з'являється можливість інтегральної оцінки поглядів за допомогою різних норм, а також близькості різних поглядів. За певних умов маємо евклідовий простір, в якому можна визначити поняття ортогональності (незалежності) поглядів. Можуть бути використані й інші норми, що найбільш точно відображають зміст «близькості» поглядів.

Для оцінювання «близькості» статистичних розподілів скористаємось відомою мірою $L(P, Q) = 2H((P \oplus Q)/2) - H(P) - H(Q)$, де P і Q – оцінювані розподіли, $(P \oplus Q)/2 = \langle (p_1 + q_1)/2, (p_2 + q_2)/2, \dots, (p_n + q_n)/2 \rangle$ – усереднений розподіл, $H((P \oplus Q)/2)$, $H(P)$ і $H(Q)$ – ентропія Шеннона. Виходячи з потреб вирішення конкретної прикладної задачі, у роботі ця міра модифікована з використанням найкращого і найгіршого розподілів і покладена в основу оцінювання при узгодженні даних у вигляді формули

$$S(P) = (L(P, P_{inf}) - L(P_{inf}, P_{inf})) / (L(P_{sup}, P_{inf}) - L(P_{inf}, P_{inf})) \quad (7)$$

Оскільки $L(P_{inf}, P_{inf}) = 0$, формула (7) набуває вигляду $S(P) = L(P, P_{inf}) / L(P_{sup}, P_{inf})$, при цьому $S(P) \in \{0, 1\}$. Отримане значення прийняте за кількісну оцінку розподілу, виражену у відносній шкалі. На множині розподілів $\pi = \{P_j, j = \overline{1, m}\}$ визначений нестрогий порядок $\alpha \subseteq \pi \times \pi: \langle p_i, p_j \rangle \in \alpha \Leftrightarrow S(p_i) \geq S(p_j)$ з границями P_{sup} і P_{inf} . Бажаний критерій надає вирішення оптимізаційних задач вигляду $\max_{P \in \pi} (S(P)), \arg\max_{P \in \pi} (S(P))$.

На основі виконаних досліджень втрат і метрик у роботі запропоновано метод перетворення даних, який забезпечує відсутність інформаційних втрат, якщо типи вихідної і цільової шкали збігаються, і мінімізує втрати у іншому випадку.

Представлені в кількісних шкалах $X^j \in R, j = \overline{1, m}$, де R – поле дійсних чисел, дані приводяться до єдиної семантики і діапазону $C_j: X^j \xrightarrow{C_j} R$ без інформаційних втрат, що накладає умова монотонності $x_{k,j} \prec x_{l,j} \Leftrightarrow C_j(x_{k,j}) \leq C_j(x_{l,j}); k, l \in [1, n], k \neq l$, за допомогою параметричних перетворень на основі лінійних $C_{norm}(x_{i,j}) = \frac{x_{i,j} - a}{b}$ і нелінійних

$C_{norm}(x_{i,j}) = \left(1 - e^{\frac{a - x_{i,j}}{b}}\right)^{-1}$ нормувань. Тут a, b – параметри зміщення і масштабу нормування, для обчислення значень яких використовуються вирази $a = \min_{i=1, n}(x_{i,j}), b = \max_{i=1, n}(x_{i,j}) - \min_{i=1, n}(x_{i,j})$ і $a = \overline{X^j} = \frac{1}{n} \sum_{i=1}^n x_{i,j}, b = \sigma(X^j) = \sqrt{\left(\sum_{i=1}^n (x_{i,j} - \overline{X^j})^2\right) / n}$, де $\overline{X^j}$ – середнє значення, а $\sigma(X^j)$ – стандартне відхилення показника X^j , що визначається на вибірці об'єктів O (нормується середнє по вибірці O).

Якщо за еталони приймаються об'єкти, які характеризуються граничними станами, то відображення оцінювання інтерпретуватися як функції приналежності для лінгвістичних змінних. Якщо показники задані в номінальних шкалах, спектр еквівалентних перетворень даних обмежується бієкціями. У роботі розроблена схема бієктивного узгодження даних, представлених в номінальних шкалах і в рамках якої завдання узгодження даних розглядається як задача оптимізації:

- Крок 1. Актуалізація $O, N, M, X \subseteq O \times N, Y \subseteq O \times M$;
- Крок 2. Подання функцій $b: N \rightarrow M, b^{-1}: M \rightarrow N$;
- Крок 3. Визначення $\forall_{n \in N} b^{-1}(b(n)) = n, \forall_{m \in M} b(b^{-1}(m)) = m$;
- Крок 4. Визначення $Y \subseteq O \times b^{-1}(M)$;
- Крок 5. Знаходження $\arg\max_{b \in B} a_k(X, Y')$.

За критерій оптимізації у цій схемі прийнята ймовірнісна міра – альфа Кліппендорфа $\alpha_k = ((n-1) \sum_{i=1}^n o_{i,i} - \sum_{i=1}^n s_i(s_i - 1)) / (n(n-1) - \sum_{i=1}^n s_i(s_i - 1))$.

Значення α_k в діапазоні 0,75-1,0 відповідають високому ступеню, 0,5-0,75 – середньому ступеню, а значення менше 0,5 – низькому ступеню узгодженості даних, представлених в номінальних шкалах.

Для перетворення з мінімальними втратами даних, представлених в різних шкалах, запропоновано кілька методів. Так, для перетворення в порядкову шкалу даних, представлених в кількісних шкалах, для мінімізації інформаційних втрат запропонований метод, що базується на побудові варіаційного ряду $\tilde{X} = \langle \tilde{x} \rangle : \forall \tilde{x} \in \tilde{X}, \tilde{x} \in X, \forall \tilde{x}_i, \tilde{x}_j \in \tilde{X}, i < j \quad \tilde{x}_i \leq \tilde{x}_j$ для показника X , вираженого в кількісній шкалі, та кусочно-лінійній апроксимації його кумуляти, яка визначається наступним чином: $C_k = \langle c_i \rangle : c_i = \sum_{l=1}^i \tilde{x}_l, i = \overline{1, n}$, де c_i – i -те значення кумуляти, $\tilde{x}_l \in \tilde{X}$ – l -тий член варіаційного ряду, n – довжина цього ряду.

Визначення значень показника X в порядковій шкалі пов'язане з побудовою розбиття $\pi(\tilde{X}) = \langle \langle \tilde{x}_1, \tilde{x}_2 \rangle, \langle \tilde{x}_2, \tilde{x}_3 \rangle, \dots, \langle \tilde{x}_{r-1}, \tilde{x}_r \rangle \rangle, \tilde{x}_l \in \tilde{X}, l = \overline{1, r}, \forall l_1, l_2 = \overline{1, r}, l_1 < l_2 : \tilde{x}_{l_1} \leq \tilde{x}_{l_2}$, для якого сума квадратів відхилень значень показника X сегментів $\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle \in \pi(\tilde{X})$ варіаційного ряду від математичного очікування значень показника $M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle)$ в цих сегментах варіаційного ряду є найменшою. Таким чином, кожному сегменту $\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle \in \pi(\tilde{X})$ ставимо у відповідність величину $M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle)$, а кожній з них, у свою чергу, – значення ординальної шкали $X \rightarrow \tilde{X} \rightarrow \pi(\tilde{X}) \rightarrow \langle M(\langle \tilde{x}_i, \tilde{x}_{i+1} \rangle) \rangle, i = \overline{1, r-1} \rightarrow \overline{1, r-2}$.

В четвертому розділі розглядаються технологічні аспекти реалізації платформи підтримки міждисциплінарних досліджень в Світовій системі даних.

Розглянуті у попередніх розділах концепція, моделі, методи та алгоритми становлять складові теоретичних основ реалізації зазначеної платформи. Платформа підтримки міждисциплінарних досліджень в Світовій системі даних у якості базових складових включає програмний інструментарій і комплекс методик його застосування для створення інформаційних систем ССД, їх експлуатації, побудови запитів, участі у формуванні схем розв'язання проблем, оцінювання результатів. Сервіс-орієнтована архітектура платформи наведена на рис. 5.

Призначення платформи в цілому визначає функціональність окремих підсистем та архітектурних блоків, зображених на рисунку. Основними архітектурними компонентами платформи є:

1. Підсистема довготривалого зберігання даних, яка має функціональність, необхідну для вирішення задач управління даними.
2. Підсистема опрацювання даних, яка реалізує алгоритми попередньої обробки, верифікації та аналітичного опрацювання даних.
3. Підсистема підтримки прикладних застосунків, що дозволяє вирішувати задачі розробки прикладних застосунків та управління ними.
4. Загальний реєстр ресурсів, який дозволяє управляти конфігурацією розподіленої системи та здійснювати моніторинг її стану.

Розроблена платформа є основою розподіленої інформаційної системи з сервіс-орієнтованою архітектурою, призначеною для вирішення задач, що виникають в

процесі реалізації безперервного життєвого циклу даних, які використовуються для наукових та прикладних досліджень. Платформа містить ресурси, що забезпечують моніторинг стану індикаторів, має розвинуті механізми аналізу даних та візуалізації. Набір засобів аналітичної обробки даних створено на основі сучасних інформаційних технологій, що забезпечує їх високу надійність та зручність у використанні. Кожен інструмент створено у вигляді окремого сервісу, що відповідає за роботу того чи іншого алгоритму обробки даних.

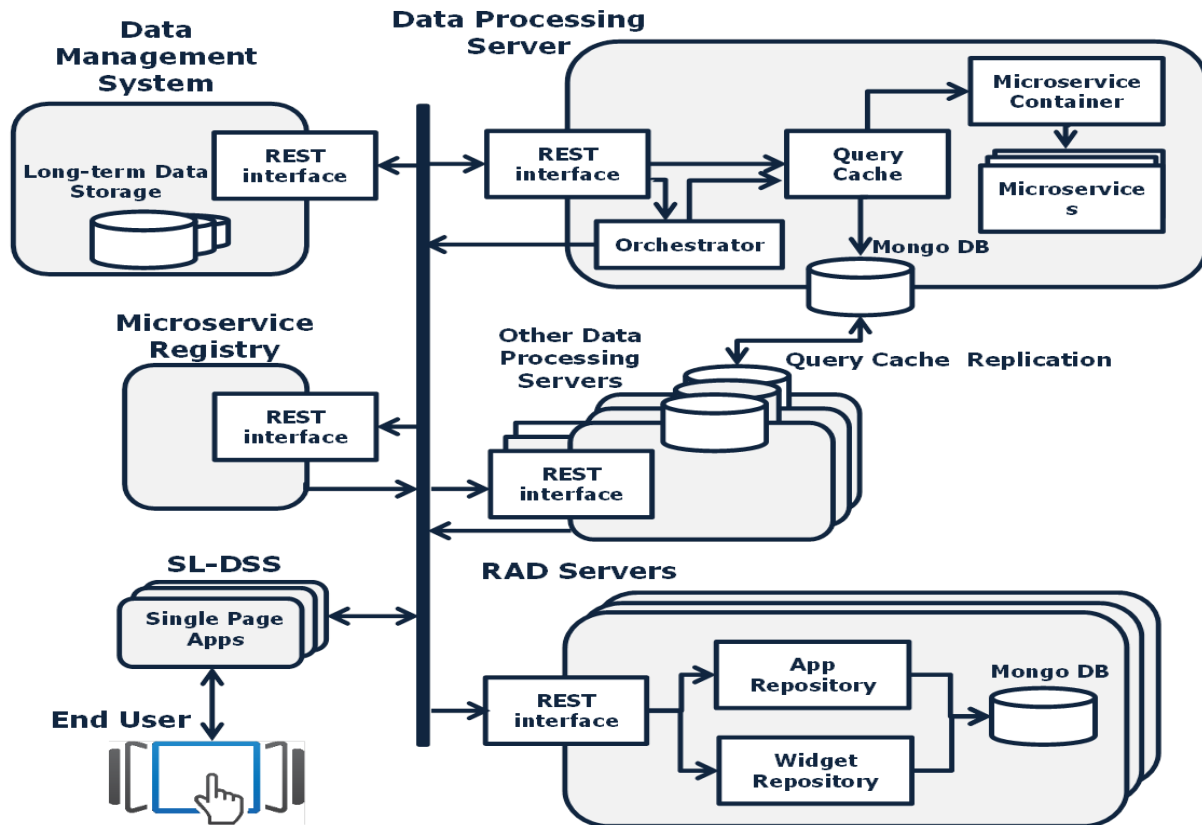


Рисунок 5 – Сервіс-орієнтована архітектура платформи

Дані в підсистемах довготривалого зберігання та опрацювання подаються у вигляді наборів (у форматі JSON та інших), що описують багатовимірні куби та супроводжуються метаданими, необхідними для визначення семантики цих даних. Кінцевий користувач має змогу описувати дані у вигляді файлу в форматі .xlsx, який складається з аркушів, призначених для опису метаданих набору в цілому, власне даних, розширення метаданих (тлумачного словника) та засобів інтернаціоналізації. Кожен з наборів ідентифікується за допомогою універсального унікального ідентифікатора (UUID), що забезпечує можливість міграції наборів даних між окремими вузлами. Ця підсистема надає можливість контролю версій даних завдяки збереженню управління ланцюжками версій, які супроводжуються коментарями адміністратора даних або генерованих засобами автоматичного оновлення даних повідомленнями. Метадані, що супроводжують набір, подаються деревоподібною ієрархічною структурою визначеної форми.

Підсистема підтримки прикладних застосунків надає користувачеві набір інструментів, необхідний для створення інтерактивних багатосторінкових звітів з результатами досліджень на основі аналітичного опрацювання даних. Вона

відповідає жорсткому комплексу вимог: online-розробка і хостинг односторінкових застосунків; персональні тематичні панелі на основі використання віджетів; взаємодія віджетів; гнучка схема навігації і компоновка сторінок; стилі і повна інтернаціоналізація застосунків; управління ресурсами застосунків; підтримка взаємодії користувачів; авторизація доступу до ресурсів та ін.

У рамках підсистеми реалізовані набори інструментів, які призначені для: управління застосунками; управління даними (підтримка метаданих, повна інтернаціоналізація наборів даних, контроль версій даних, імпорт/експорт даних); інтелектуального опрацювання та візуалізації даних (маніпулювання багатовимірними кубами даних, запити проєкції, вибірки, злиття, інтерактивні графіки, багат шарові тематичні мапи, таблиці, вбудовані інструменти опрацювання та аналізу даних); аналізу часових послідовностей подій.

У роботі можливості платформи продемонстровані на прикладі створення застосунку, призначеного для побудови стратегій та сценаріїв сталого розвитку України (доступ за посиланням <http://open.wdc.org.ua/>). Створений застосунок складається з низки інтерактивних тематичних панелей (dashboard) та дозволяє в зручній для особи, що приймає рішення, формі здійснити моделювання та порівняльний аналіз показників сталого розвитку України в двох розрізах:

1. В глобальному розрізі, коли індикатори сталого розвитку України порівнюються з відповідними індикаторами інших країн Європи і світу;
2. В регіональному розрізі, коли проводиться моделювання і порівняльний аналіз сталого розвитку регіонів (областей) України.

Результати моделювання і аналізу використовуються для оцінювання наближення/віддалення соціально-економічного розвитку України до бажаних сценаріїв «Оптимістичного» (до 2020 р.) і «Збалансований розвиток» (до 2030 р.).

У **висновках** сформульовано основні результати дисертаційної роботи.

Розроблені на базі платформи програмні інструменти для інтелектуальної обробки, аналізу та системного узгодження даних різної природи, їх систематизації, оцінювання, аналізу якості, коректності, продемонстровано на прикладі вирішення таких прикладних задач:

1. Моделювання та кількісне оцінювання процесів сталого розвитку територій (регіонів України, країн) в контексті якості та безпеки життя людей і оцінювання ступеня впливу сукупності загроз на їх розвиток (**Додаток А**).
2. Оцінювання ефективності масових суспільних заходів (перетворень), зокрема оцінювання ставлення суспільства до проведення масової вакцинації проти COVID-19 в Україні (**Додаток Б**).

Документи, які підтверджують результати впровадження, наведено в **Додатку В**.

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну науково-практичну задачу організації, планування і здійснення підтримки міждисциплінарних досліджень в ССД. Проведені дослідження дозволили отримати наступні наукові результати, які мають істотні переваги перед існуючими рішеннями:

1. Виконано аналіз діяльності Світової системи даних і окремого Світового центру даних як об'єкта управління, виявлено і проаналізовано невирішені проблеми підтримки міждисциплінарних досліджень, обґрунтовано перспективність створення платформи підтримки міждисциплінарних досліджень, як універсального інструментарію вирішення проблем організації діяльності Світової системи даних.

2. *Вперше створено* платформу підтримки міждисциплінарних досліджень в Світовій системі даних, яка забезпечує інтеграцію застосунків і джерел даних, оброблення запитів до розподілених джерел, системне узгодження даних різної природи, проблемно-орієнтоване інтелектуальне оброблення даних та автоматизоване генерування застосунків.

3. *Вперше створено* метод узгодження даних, який відрізняється застосуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами.

4. *Модифіковано* клаузальну логіку планування взаємодії застосунків для розв'язання проблем міждисциплінарних досліджень, яка відрізняється від відомих використанням системи типів, орієнтованих на організацію взаємодії застосунків, комплексом спеціальних аксіом описання застосунків на основі передумов та постулатів і правил виведення, які визначають принципи комплексування застосунків, що дозволяє у зручному для користувача вигляді визначати проблему і будувати схему її вирішення.

5. *Отримав подальший розвиток* метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в просторі типів з наступною деталізацією виводу в системі індивідних об'єктів та відновленням схеми розв'язання проблеми на основі виводу.

6. Компоненти розробленої платформи підтримки міждисциплінарних досліджень в Світовій системі даних знайшли впровадження в Інституті проблем реєстрації інформації НАН України, Інституті геофізики НАН України ім. С.І. Субботіна, Інформаційно-аналітичному ситуаційному центрі КПІ ім. Ігоря Сікорського, Навчально-науковому комплексі «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського, Корпорації «Науковий парк «Київська Політехніка» та інших організаціях, що дозволило підтвердити їх працездатність, скоротити проектування і реалізацію застосунків на 20 – 30 % в залежності від особливостей діяльності, скоротити втрати внаслідок невчасного отримання потрібних даних на 11,75 %, скоротити інформаційні втрати при перетворенні даних на 7,25%, підвищити обсяги наданих користувачам інформаційних ресурсів на 50%, підвищити ефективність аналізу даних різної природи експертами без знання інструментальних мов програмування на 42%.

7. Розроблені методичні та програмні засоби використовуються на кафедрі математичних методів системного аналізу та кафедрі автоматики та управління в технічних системах КПІ ім. Ігоря Сікорського при підготовці фахівців в галузі інтелектуального аналізу даних та інформаційних технологій.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Монографії

1. Zgurovsky M., Boldak A., Melnyk O., Perestyuk M., Putrenko V., Pyshnograiev I., Yasinsky V., Yefremov K., Foresight 2018: systemic world conflicts and global forecast for XXI century / International Council for Science etc.; Scientific Supervisor M. Zgurovsky. — К. : NTUU «Igor Sikorsky Kyiv Polytechnic Institute», 2018. — 226 p. *Здобувачем запропоновано та розроблено мікросервіси для аналізу та прогнозування світових конфліктів; реалізацію розроблених мікросервісів імplementовано до інтегрованої он-лайн платформи Advanced Analytics ННК «Світовий центр даних з геоінформатики та сталого розвитку».*
2. Згуровський М.З., Болдак А.О., Джигирей І.М., Єфремов К.В., Форсайт 2018: Аналіз підготовки і перепідготовки фахівців природничого і технічного спрямування, виходячи з цілей сталого соціально-економічного розвитку України до 2025 року. — К. : НТУУ «КПІ імені Ігоря Сікорського», Вид-во «Політехніка», 2018. — 32 с. *Здобувачем створено та реалізовано засобами розробленої платформи підтримки міждисциплінарних досліджень сценарій (ланцюг) послідовного та паралельного виконання мікросервісів для виконання аналізу підготовки і перепідготовки фахівців, запропоновано і реалізовано додаткові необхідні мікросервіси в рамках зазначеної платформи. Результати дослідження представлено за допомогою інструментів розробленої платформи.*
3. Foresight and construction of the strategies of socio-economic development of Ukraine on mid-term (up to 2020) and long-term (up to 2030) time horizons / Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky // International Council for Science (ICSU); Committee for the System Analysis of the Presidium of NAS of Ukraine; National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»; Institute for Applied System Analysis of MES of Ukraine and NAS of Ukraine; World Data Center for Geoinformatics and Sustainable Development; Agrarian Superstate Foundation. — 2nd ed. — Kyiv : NTUU «Igor Sikorsky KPI», Publ. house «Polytechnica», 2016. — 184 p. *Здобувачем розроблено застосунки на основі мікросервісів платформи підтримки міждисциплінарних досліджень для проведення сценарного моделювання поведінки складних соціально-економічних систем; виконано інтеграцію необхідних джерел даних, створено віджети для відображення результатів моделювання.*
4. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — К. : Igor Sikorsky KPI, 2017. — Part 2. Ukraine in Sustainable Development Indicators (2016–2017). — 72 p. *Здобувачем запропоновано та реалізовано процедури розрахунку індексу сталого розвитку країн світу, компоненти безпеки життя, компоненти якості життя та її складових з використанням логічного підходу для планування взаємодії мікросервісів платформи підтримки міждисциплінарних досліджень. Здобувачем виконано опис та інтеграцію (засобами платформи) інформаційних ресурсів, необхідних для дослідження.*

5. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — K. : Igor Sikorsky KPI, 2017. — Part 1. Global Analysis of Quality and Security of Life (2016). — 208 p. *Здобувачем запропоновано та реалізовано процедури розрахунку індексу сталого розвитку регіонів України, компоненти безпеки життя, компоненти якості життя та її складових з використанням логічного підходу для планування взаємодії мікросервісів платформи підтримки міждисциплінарних досліджень. Здобувачем виконано опис та інтеграцію (засобами платформи) інформаційних ресурсів, необхідних для дослідження.*

Стаття у періодичному науковому виданні іншої держави, яка входить до ОЕСР та Європейського Союзу

6. Marsel Shaimardanov, Alexei Gvishiani, Michael Zgurovsky, Alexander Sterin, Alexander Kuznetsov, Natalia Sergeyeva, Evgeny Kharin, Kostiantyn Yefremov. Development of WDS Russian-Ukrainian segment // Data Science Journal. — 2013. — Volume 12. — p. 17-26. (журнали країн, що **входять до ОЕСР, ЄС**; включене до бази даних **Scopus**). *Здобувачем описано і проаналізовано типовий процес інтелектуальної обробки даних в міждисциплінарних дослідженнях, сформульовано вимоги до моделей і методів інтелектуальної обробки даних в міждисциплінарних дослідженнях, запропоновано архітектурні принципи організації системи інтелектуальної обробки даних в рамках інформаційно-комунікаційної інфраструктури Світового центру даних з геоінформатики та сталого розвитку і його партнерів.*

Статті у наукових фахових виданнях

7. M.Z. Zgurovsky, A.A. Boldak, K.V. Yefremov. Intelligent analysis and the systemic adjustment of scientific data in interdisciplinary research // Cybernetics and Systems Analysis: Volume 49, Issue 4 (2013), pp. 541-552. (журнал **категорії «А»**, включене до бази даних **Scopus**). *Здобувачем виконано вербальний опис задачі узгодження даних в міждисциплінарних дослідженнях, запропоновано формальні моделі задачі узгодження і розроблено метод узгодження даних з врахуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, отриманих з різних джерел, для конструювання процедури перетворення даних з найменшими інформаційними втратами.*
8. Теленик С.Ф., Амонс О.А., Єфремов К.В., Лиско В.Т. Логічний підхід до інтеграції програмних застосувань підтримки міждисциплінарних наукових досліджень // Наукові вісті НТУУ «КПІ». — 2013. — № 5. — С. 53-72. *Здобувачем в рамках логічного підходу до інтеграції застосунків запропоновано модифікацію формальної логічної системи на основі клаузальної логіки першого порядку для планування взаємодії застосунків, удосконалений з врахуванням дворівневого характеру логічного формалізму метод виведення для формальної логічної системи і механізм побудови схеми виконання запиту користувача на основі виводу.*
9. Теленик С.Ф., Амонс О.А., Єфремов К.В., Жук С.В. Семантична інтеграція різнорідних інформаційних ресурсів // Вісник НТУУ «КПІ». Інформатика,

управління та обчислювальна техніка: Зб. наук. пр. – К.: Век+, – 2013. – No. 58. – с.29-45. *Здобувачем запропоновано загальний підхід до вирішення проблеми інтеграції інформаційних ресурсів Світової системи даних на основі web-орієнтованих технологій, способу організації взаємодії джерел даних з посередником і моделей дескриптивної логіки в термінах OWL, RDF і SPARQL, а також схему оброблення запитів до розподілених джерел даних, моделі і методи формування плану виконання запиту та його оптимізації, формування запитів безпосередньо до джерел даних, алгоритм побудови зв'язаної структури даних із онтологій та підключень до джерел даних.*

10. Згуровский М.З., Болдак А.А., Ефремов К.В., Сергеева Н.А., Забаринская Л.П., Шестопапов И.П., Нисилевич М.В. Применение методов интеллектуального анализа данных для эмпирических исследований взаимосвязи гелио- и геофизических процессов // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: Век+, – 2013. – No. 58. – с.4-10. *Здобувачем розроблено на основі логічного підходу і реалізовано засобами платформи підтримки міждисциплінарних досліджень сценарії (сукупність взаємопов'язаних процесів) обробки даних для вирішення задачі аналізу взаємозв'язку геліо- і геофізичних процесів.*
11. Болдак А.О., Єфремов К.В. Предметно-орієнтована мова аналітичної обробки даних // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: Век+, – 2012. – No. 55. – с.67-71. *Здобувачем запропоновано загальний підхід до розроблення і надання користувачам Світової системи даних нових інформаційних ресурсів і проблемно-орієнтовану графічну мову аналітичної обробки даних як основу швидкого розроблення застосунків на платформі підтримки міждисциплінарних досліджень.*
12. M. Z. Zgurovsky, A. D. Gvishiani, K. V. Yefremov and A. M. Pasichny. Integration of the Ukrainian science into the world data system // Cybernetics and Systems Analysis: Volume 46, Issue 2 (2010), pp. 211-219. (журнал **категорії «А»**, включене до бази даних **Scopus**). *Здобувачем виконано аналіз діяльності Світової системи даних і Світового центру даних з геоінформатики та сталого розвитку як об'єкта управління, виявлено і проаналізовано невирішені проблеми підтримки міждисциплінарних досліджень в Світовій системі даних, сформульовано принципи формування ІТ-середовища, здатного забезпечити ефективну підтримку міждисциплінарних досліджень і формування на цій основі нової ІТ-спільноти.*

Матеріали науково-технічних конференцій

13. M. Zgurovsky, A. Boldak, D. Lande, K. Yefremov and M. Perestyuk. Predictive Online Analysis of Social Transformations based on the Assessment of Dissimilarities between Government Actions and Society's Expectations. Conference proceedings of 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2020, P. 130-135. (включене до бази даних **Scopus**). *Здобувачем на базі інтегрованої он-лайн платформи Advanced Analytics Світового центру даних з геоінформатики та сталого розвитку розроблено сукупність інструментів та сервісів для кількісного оцінювання параметрів і характеристик суспільних*

перетворень (на прикладі кількісного оцінювання ставлення населення України до дій влади, пов'язаних з посиленням карантинних заходів, спрямованих на протидію поширенню пандемії COVID-19).

14. Nowakowski, G., Telenyk, S., Yefremov, K., Khmeliuk, V. The approach to applications integration for world data center interdisciplinary scientific investigations. Proceedings of the 2019 Federated Conference on Computer Science and Information Systems. FedCSIS 2019, P. 539-545. (включене до бази даних **Scopus**). *Здобувачеві належить архітектура системи інтеграції застосунків в Світовій системі даних для підтримки міждисциплінарних досліджень на основі оркестрації як способу організації взаємодії сервісів, механізм якої базується на методах представлення семантики перетворень, виконуваних сервісами, методах виведення в клаузальній логіці і методах відновлення схем виконання запитів користувачів на основі виводу.*
15. Zgurovsky, M., Putrenko, V., Dzhygyrey, I., Boldak, A., Yefremov, K., Pashynska, N., Pyshnograiev, I., Nazarenko, S. Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine. Conference proceedings of 2018 IEEE 1st International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2018, P. 1-5. (включене до бази даних **Scopus**). *Здобувачем запропоновано загальний підхід до оцінювання сталого розвитку територій на основі яскравості нічних вогнів; на базі інтегрованої он-лайн платформи *Advanced Analytics* Світового центру даних з геоінформатики та сталого розвитку розроблено сукупність інструментів та сервісів для встановлення залежностей між супутниковими даними яскравості нічних вогнів та показниками сталого розвитку на прикладі території України.*
16. Zgurovsky, M., Boldak, A., Yefremov, K., Pyshnograiev, I. Modeling and investigating the behavior of complex socio-economic systems. 2017. IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 – Proceedings, P. 1113-1116. (включене до бази даних **Scopus**). *Здобувачем запропоновано реалізацію серії методів для оцінки поточного стану складних соціально-економічних систем, прогнозування їх майбутніх станів; розроблено та реалізовано засобами платформи підтримки міждисциплінарних досліджень сценарії (сукупність взаємопов'язаних мікросервісів) обробки даних для дослідження складних соціально-економічних систем; представлено за допомогою інструментів розробленої платформи результати дослідження.*
17. Telenyk, S., Nowakowski, G., Yefremov, K., Khmeliuk, V. Logics based application integration for interdisciplinary scientific investigations. Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017, P. 1026-1031. (включене до бази даних **Scopus**). *Здобувачеві належить підхід до інтеграції сервісів програмних систем Світової системи даних на основі автоматизованої побудови схеми взаємодії сервісів для виконання запиту користувача шляхом виведення в клаузальній логіці, логічні моделі опису сервісів (застосунків) та джерел даних, методи виведення та відновлення схеми виконання запиту на основі виводу, вимоги до складу і взаємодії компонентів підтримки міждисциплінарних досліджень в Світовій системі даних з урахуванням стану ІТ-галузі.*

АНОТАЦІЯ

Єфремов К.В. Платформа підтримки міждисциплінарних досліджень в Світовій системі даних. На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології. Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Міністерство освіти і науки України, Київ, 2021.

Дисертаційну роботу присвячено вирішенню важливої науково-практичної задачі організації, планування і здійснення підтримки міждисциплінарних досліджень в Світовій системі даних. Розроблено концепцію платформи підтримки міждисциплінарних досліджень в ССД. Вперше створено метод узгодження даних різної природи з врахуванням оцінок інформаційних втрат процедур перетворення даних та оцінок узгодженості даних, модифіковано клаузальну логіку планування взаємодії застосунків для розв'язання проблем міждисциплінарних досліджень, отримав подальший розвиток метод виведення в клаузальній логіці взаємодії застосунків за рахунок попереднього виведення в просторі типів з наступною деталізацією виводу в системі індивідних об'єктів та відновленням схеми розв'язання проблеми на основі виводу. На основі розробленої концепції і оригінального комплексу моделей, методів і алгоритмів реалізовано інструментальні засоби платформи підтримки міждисциплінарних досліджень в ССД, яка забезпечує інтеграцію застосунків і джерел даних, оброблення запитів до розподілених джерел, системне узгодження даних різної природи, проблемно-орієнтоване інтелектуальне оброблення даних та автоматизоване генерування застосунків. Проведено експериментальне дослідження створеної платформи на прикладі вирішення низки прикладних задач, у т.ч. аналізу та прогнозування світових конфліктів, побудови стратегій та сценаріїв сталого розвитку України, аналізу і прогнозного моделювання поширення епідемії COVID-19 в Україні та ін. Платформа та окремі її компоненти успішно використовуються в низці організацій.

Ключові слова: Світова система даних, Світовий центр даних, інтеграція застосунків, інтеграція джерел даних, клаузальна логіка, системне узгодження даних, міждисциплінарні дослідження.

АННОТАЦИЯ

Ефремов К.В. Платформа поддержки междисциплинарных исследований в Мировой системе данных. На правах рукописи.

Диссертация на соискание учёной степени кандидата технических наук по специальности 05.13.06 — Информационные технологии. Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Министерство образования и науки Украины, Киев, 2021.

Диссертационная работа посвящена решению важной научно-практической задачи организации, планирования и осуществления поддержки междисциплинарных исследований в Мировой системе данных. Разработана концепция платформы поддержки междисциплинарных исследований в МСД. Впервые создан метод согласования данных разной природы с учетом оценок информационных потерь

процедур преобразования данных и оценок согласованности данных, модифицирована клаузная логика планирования взаимодействия приложений для решения проблем междисциплинарных исследований, получил дальнейшее развитие метод вывода в клаузной логике взаимодействия приложений за счет предварительного вывода в пространстве типов с последующей детализацией вывода в системе индивидуальных объектов и восстановлением схемы решения проблемы на основе вывода. На основе разработанной концепции и оригинального комплекса моделей, методов и алгоритмов реализованы инструментальные средства платформы поддержки междисциплинарных исследований в МСД, которая обеспечивает интеграцию приложений и источников данных, обработку запросов к распределенным источникам, системное согласование данных разной природы, проблемно-ориентированную интеллектуальную обработку данных и автоматизированную генерацию приложений. Проведено экспериментальное исследование созданной платформы на примере решения ряда прикладных задач, в т.ч. анализа и прогнозирования мировых конфликтов, построения стратегий и сценариев устойчивого развития Украины, анализа и прогнозного моделирования распространения эпидемии COVID-19 в Украине и др. Платформа и отдельные ее компоненты успешно используются в ряде организаций.

Ключевые слова: Мировая система данных, Мировой центр данных, интеграция приложений, интеграция источников данных, клаузная логика, системное согласование данных, междисциплинарные исследования.

ABSTRACT

Yefremov K.V. A platform to support interdisciplinary research in the World Data System. As a manuscript.

Thesis for a candidate degree in technical sciences, specialty 05.13.06 – Information Technologies. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ministry of Education and Science of Ukraine, Kyiv, 2021.

The dissertation is devoted to solving an important scientific and practical problem – improving the efficiency of information and software resources of the World Data System, developing approaches and implementing tools to support interdisciplinary research in the World Data System.

Members of the World Data System since its inception in 1956 have accumulated a large amount of data and many applications (information systems, services), which are in great demand in international research aimed at finding answers to global challenges facing humanity. However, this is hindered by difficulties in the interaction of legacy applications, the use of different formats and methods of data access, and so on. Solving the problems of analysis, forecasting, scenario modeling of crisis and security phenomena and studying their impact on the economy and society, such as global threats to sustainable development, spreading the SARS-CoV-2 pandemic, etc., also require tools for rapid integration of data from different sources and applications. The problem of organizing, planning, and supporting interdisciplinary research in WDS arises. Varying degrees of depth on particular aspects of integration has led to the need for effective models, methods, and tools needed to implement these solutions.

Within the framework of the dissertation, an analysis of the World Data System and a separate World Data Center as a controlled object was performed, unresolved issues of support for interdisciplinary research were identified and analyzed, the prospects of creating the platform to support interdisciplinary research were substantiated as a universal tool for solving problems of organizing the World Data System functioning. The concept of the platform to support interdisciplinary research in the World Data System was developed. For the first time, a method was proposed for the harmonization of data of different nature, taking into account estimates of information loss of data conversion procedures and estimates of consistency of data obtained from different sources, to design a procedure for converting data with the lowest information loss. The dissertation proposes the MAS architecture for the transformation of heterogeneous but semantically identical information into metadata, suitable for further analysis and research. The clausal logic of planning the interaction of applications to solve problems of interdisciplinary research was modified. A formal logical system was developed. This formalism is the basis of the integration component of the platform is built. The components of the formal system were described: the language of clausal first-order logic; knowledge base, which is based on an ontology of axioms depicting the methods of services, and an ontology of data sources described in the OWL language based on RDF; inference rules necessary to obtain the desired result in cases where it is necessary to combine methods of both one and different services. The inference method in the clausal logic of application interactions was further developed. The proposed inference method is based on the typification of statements and analogies, it is the formal basis of the inference mechanism in the platform. The output constructed in abstract space is used to control the inference process in the initial solution search space, which will increase the inference efficiency by cutting off most of the unpromising inference branches in the initial space. The algorithm of the solution scheme recovery for the services (applications) integration was developed as a basis for the mechanism of the solution scheme recovery to obtain the functional sequence of actions (scenario) from the output of the inference mechanism. The scenario is executed by the system, taking into account the features and purpose of services (applications). The service-oriented architecture of the platform to support interdisciplinary research was developed, the principles of functioning of interaction of its components were described.

The tools of the platform to support interdisciplinary research in the World Data System were developed based on the proposed concept and original set of models, methods, and algorithms. The Platform provides integration of applications and data sources, processing queries to distributed sources, systemic harmonization of data of various nature, problem-oriented intelligent data processing and automated application generation. An experimental study of the created Platform was conducted on the example of solving a number of applied problems, including analysis and forecasting of world conflicts, construction of strategies and scenarios of sustainable development of Ukraine, analysis and forecast modeling of the COVID-19 epidemic spread in Ukraine, etc.

Components of the developed platform to support interdisciplinary research in the World Data System were implemented at the Institute for Information Recording of the National Academy of Sciences of Ukraine, the Institute of Geophysics by S.I. Subbotin name of the National Academy of Sciences of Ukraine, the Information and Analytical Situation Center of Igor Sikorsky Kyiv Polytechnic Institute, the Educational and Scientific

Complex "World Data Center for Geoinformatics and Sustainable Development" of Igor Sikorsky Kyiv Polytechnic Institute, Corporation "Science Park "Kyivska Polytechnika", Resource Efficient and Cleaner Production Centre Kyiv City Innovative Sectoral Organization of Employers, other organizations, which allowed to confirm the efficiency of the Platform, to reduce the time for applications design and implementation, to reduce losses due to late receipt of the required data, to reduce information losses in data conversion; to increase the amount of information resources provided to users, and to increase the efficiency of data analysis of various nature by experts without knowledge of instrumental programming languages

Keywords: World Data System, World Data Center, application integration, data source integration, clausal logic, systemic data harmonization, interdisciplinary research.